

AstraZeneca R&D Lund, Sweden



1/98

Nonlinear Mixed Effects Models

what they are and how they can be used

Anders Källén

Anders.Kallen@AstraZeneca.com



Presentation Overview

- Nonlinear regression, extended to generalized nonlinear regression via likelihood estimation
- The global two-stage model and the setup of a Non-Linear Mixed Effects Model
- Approximations to the likelihood of a NONMEM: First order, conditional first order, Laplace and maximum likelihood
- Where do NONMEMs come from: Population kinetics
- Some examples from my clinical experience:
 - An ACTH stimulation experiment
 - Measuring onset of action
 - A dose response experiment with a partial agonist
- Generalized linear mixed effects model, exemplified by a Poisson regression model



References

- [1] Marie Davidian and David M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Monographs on Statistics and Applied Probability 62. Chapman & Hall, first edition, 1995.
- [2] Peter J. Diggle, Kung-Yee Liang, and Scott L. Zeger. *Analysis of Longitudinal Data*. Oxford Statistical Science Series 13. Clarendon Press, Oxford, 1994.
- [3] José C. Pinheiro and Douglas M. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag New York, Inc., 2001.



Lecture 1: Nonlinear Regression

In which we look at nonlinear regression and relate it to maximum likelihood estimation. A generalization of nonlinear regression will also be introduced, accounting for variance models with auxiliary, non-regression, parameters.



Nonlinear regression: Example

Plasma concentration data of indometacin obtained after an intravenous administration:

t_i :	0.25	0.50	0.75	1.0	1.25	2.0	3.0	4.0	5.0	6.0	8.0
y_i :	2.05	1.04	0.81	0.39	0.30	0.23	0.13	0.11	0.08	0.10	0.06

To these data we want to fit the (two-compartment) regression function

$$f(\theta, t) = \theta_1 e^{-\theta_2 t} + \theta_3 e^{-\theta_4 t}.$$

The method to this is by using least squares: choose $\theta = (\theta_1, \dots, \theta_4)$ so that the sum (with $N = 11$)

$$Q(\theta) = \sum_{i=1}^N (y_i - f(\theta, t_i))^2 / \sigma_i^2$$

is minimized, where $\sigma_i^2 = V(y_i)$.

This means that we solve the equation

$$Q'(\theta) = - \sum_{i=1}^N f'(\theta, t_i)^t (y_i - f(\theta, t_i)) / \sigma_i^2 = 0.$$



Weight options:

The choices of σ_i gives weights to the residuals $y_i - f(\theta, t_i)$, and, in the absence of knowledge of the correct choice, we can use different options:

1. Ordinary Least Squares (OLS): $\sigma_i = \sigma$ for all i .
2. OLS on log-scale. Since $V(\ln y_i) \approx V(y_i)/y_i^2$, this is almost
3. $\sigma_i = y_i\sigma$, which is unnatural and should be replaced by
4. $\sigma_i = f(\theta, t_i)\sigma$, which is a special case of the power model
5. $\sigma_i = f(\theta, t_i)^\gamma\sigma$ for some parameter γ .

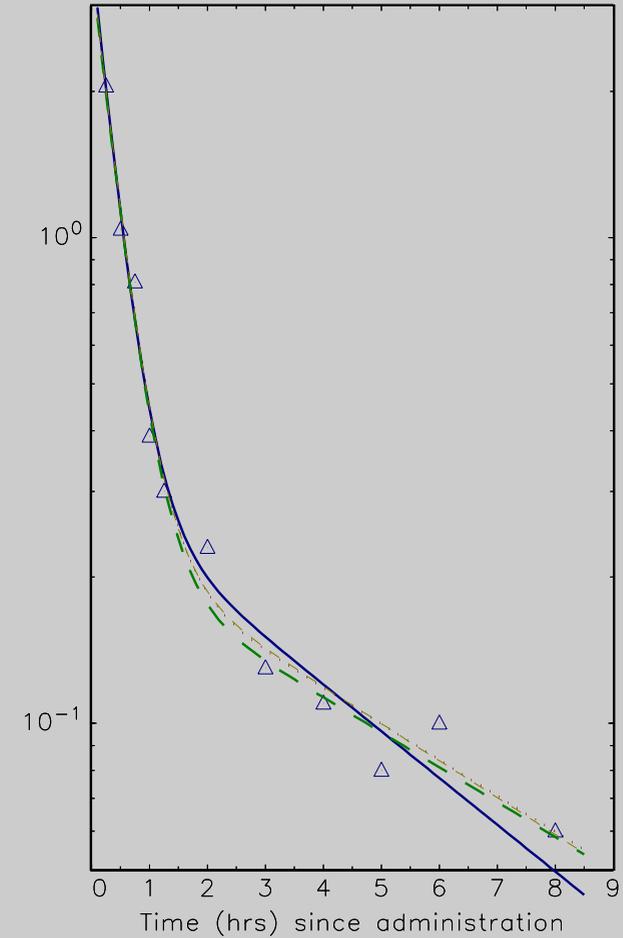
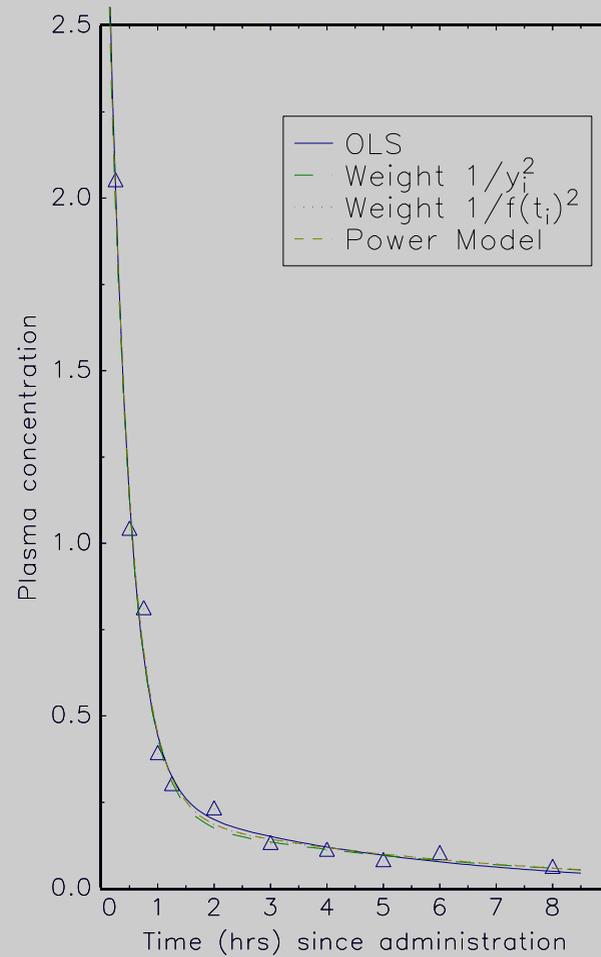
The first three of these give maximum likelihood estimates under gaussian assumptions, the last two does not – we will soon return to this.



Model fits for Indometacin data

Different weight schemes give different estimates:

	$\ln \theta_1$	$\ln \theta_2$	$\ln \theta_3$	$\ln \theta_4$
OLS	1.27	1.04	-1.23	-1.51
$1/y_i^2$	1.22	0.96	-1.52	-1.80
$1/f_i(\theta)^2$	1.21	0.95	-1.45	-1.77
$1/f_i(\theta)^{2\gamma}$	1.24	0.97	-1.43	-1.74



The estimated exponent in the power model was $\gamma = 0.82$.



Least Squares and Maximum Likelihood estimation

If we assume that the errors are independent and normally distributed,

$$y_i \in N(f(\theta, t_i), \sigma_i^2(\theta)),$$

then the likelihood

$$L(\theta) = \prod_i^N \frac{1}{\sigma_i(\theta)} \phi\left(\frac{y_i - f(\theta, t_i)}{\sigma_i(\theta)}\right), \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

implies that

$$-2 \ln L(\theta) = \sum_{i=1}^N ((y_i - f(\theta, t_i))^2 / \sigma_i(\theta)^2 + \ln(2\pi\sigma_i(\theta))).$$



To minimize this, we have to solve the equation

$$\sum_{i=1}^N f'(\theta, t_i) \sigma_i(\theta)^{-2} (y_i - f(\theta, t_i)) + \sum_{i=1}^N \sigma_i'(\theta) \sigma_i(\theta)^{-3} ((y_i - f(\theta, t_i))^2 - \sigma_i(\theta)^2) = 0.$$

This is equivalent to the least squares equation

$$\sum_{i=1}^N f'(\theta, t_i) \sigma_i(\theta)^{-2} (y_i - f(\theta, t_i)) = 0,$$

precisely when σ_i does not depend on θ !



- ML and LS estimates are different if σ_i depends on θ !
- The LS estimate focuses fully on the regression part, whereas the ML estimate is a compromise between a good fit to the regression and of the variance. Creates problems if variance model is poor!

The power model represents one further step of complication, in that $\sigma_i = \sigma_i(\theta, \gamma)$ depends on an auxiliary variance parameter.

Auxillary variance parameters are not uncommon:

- Heteroscedastic variances
- Correlation between assessments

.. leads to a variance-covariance matrix for $y = (y_1, \dots, y_n)$ like

$$V(y) = \Lambda(\theta, \phi) = U(\theta, \phi) \text{Corr}(\phi) U(\theta, \phi)^t.$$

Power model is $U(\theta, \phi) = \sigma f(\theta, t)^\phi, \quad \text{Corr}(\phi) = I.$



The Generalized Nonlinear Regression Model

This model consists of two parts:

$$E(y) = f(\theta, t), \quad V(y) = \Lambda(\theta, \phi, t), \quad (1)$$

and, under gaussian distribution assumption, its $-2 \ln L(\theta, \phi)$ is given by (matrix notations)

$$(y - f(\theta, t))^t \Lambda(\theta, \phi, t)^{-1} (y - f(\theta, t)) + \ln(\det 2\pi \Lambda(\theta, \phi, t)) \quad (2)$$

Estimating equations for ML estimation

$$\begin{cases} f'_\theta(\theta, t)^t \Lambda(\theta, \phi, t)^{-1} (y - f(\theta, t)) + \\ \quad \frac{1}{2} \lambda'_\theta(\theta, \phi)^t (\Lambda^{-1} \otimes \Lambda^{-1})(\theta, \phi, t) (s(\theta) - \lambda(\theta, \phi)) = 0 \\ \frac{1}{2} \lambda'_\phi(\theta, \phi)^t (\Lambda^{-1} \otimes \Lambda^{-1})(s(\theta) - \lambda(\theta, \phi)) = 0, \end{cases}$$

$$(s(\theta) = \text{vec}(y - f(\theta, t))(y - f(\theta, t))^t), \quad \lambda(\theta, \phi) = \text{vec}(\Lambda(\theta, \phi, t))$$

To get GLS - modify first equation and uncouple them!



Generalized Least Squares:

Iterate between

1. Solving the following equation for θ :

$$f'_{\theta}(\theta, t)^t \Lambda(\theta, \phi, t)^{-1} (y - f(\theta, t)) = 0$$

2. Minimizing the following function with respect to ϕ :

$$(y - f(\theta, t))^t \Lambda(\theta, \phi, t)^{-1} (y - f(\theta, t)) + \ln(\det 2\pi \Lambda(\theta, \phi, t))$$

until θ changes no more!

This estimation process uses ϕ only as a nuisance parameter – convergence is not necessary for it.



A soybean experiment

- This example relates to data from an experiment to compare the growth patterns of two genotypes of soybeans.
- Plots were planted and at approximative weekly intervals during the growing season, six plants were randomly selected, the leaves from these plants were aggregated and weighed, and average leaf weight per plant, y was calculated for the plot.
- To this data we want to fit a logistic growth curve

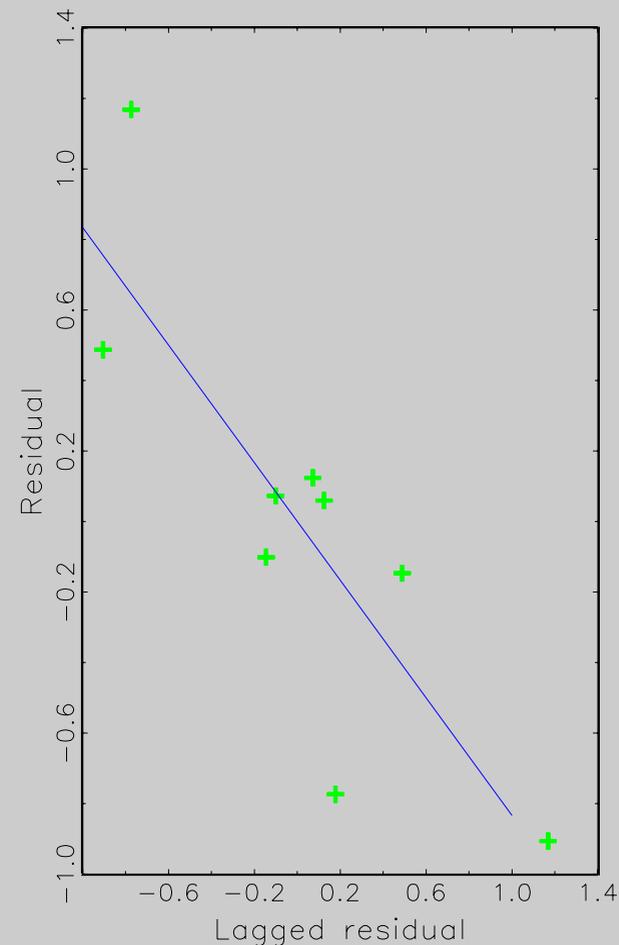
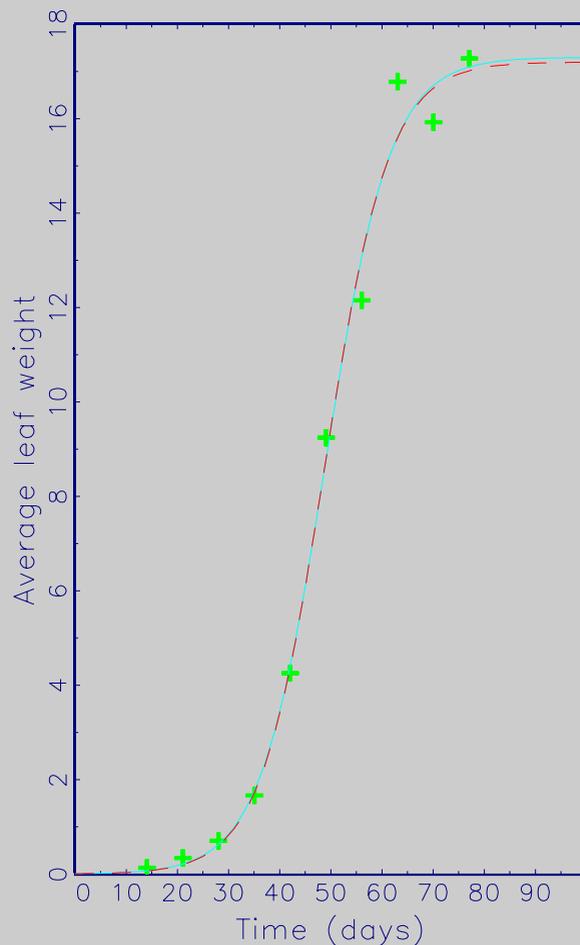
$$y(t) = \frac{\theta_1}{1 + e^{-\theta_3(t-\theta_2)}}.$$

- Here θ_1 is the asymptotic weight, θ_2 gives the time when 50% of this has been reached and θ_3 is a slope parameter. We replace θ_3 with $0.01\theta_3$ in what follows.



Analysis of the Soybean experiment

The solid curve in the plot to the left below shows the model fit. The plot to the right indicate a negative correlation (estimated to -0.83) between consecutive residuals.



The dashed curve to the right shows the analysis when a correlation structure has been imposed on the residuals in the fitting process. Virtually no difference!



The correlation model

The model we intended to use was the AR(1)-model

$$C(r_i, r_j) = \rho^{|t_i - t_j|}$$

but this is not well defined when $\rho < 0$. So, for correlation only, we recode times to consecutive integers, and use this:

$$C(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-2} & \dots & \rho & 1 & \rho \\ \rho^{n-1} & \dots & \dots & \rho & 1 \end{pmatrix}$$



So what have we gained?

Precision in parameter estimates:

First model:

Model with correlation:

Parameter	estimate	standard error	95% confidence limits	estimate	standard error	95% confidence limits
θ_1	17.3	0.5643	15.97, 18.63	17.2	0.204	16.72, 17.68
θ_2	48.85	0.8802	46.77, 50.93	48.74	0.3025	48.02, 49.45
θ_3	15.72	1.761	11.55, 19.88	15.98	0.6316	14.48, 17.47

There is some cheating in this - we assume the auxiliary parameter fixed to its estimated value (which was 0.79) when computing standard errors.



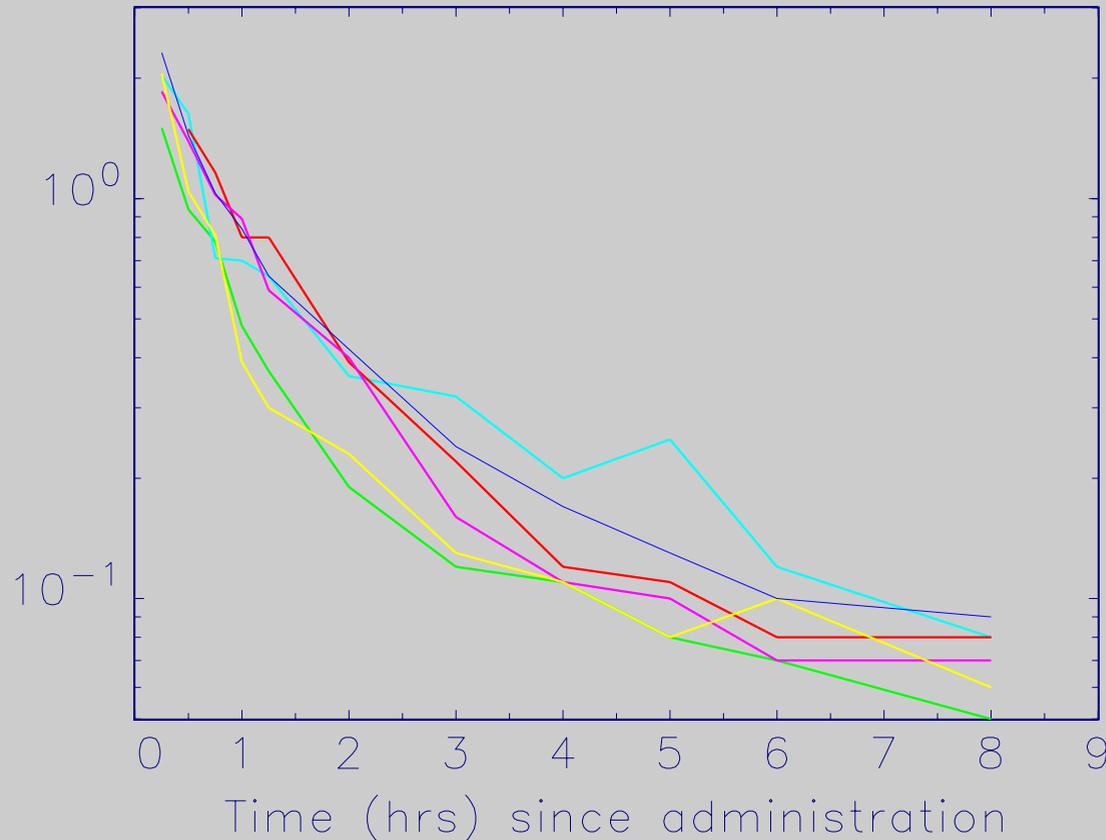
Lecture 2: The Global Two-Stage Method and the Mixed Effects Model

In which we consider the problem of summarizing the results from a sequence of nonlinear regression models into statements of how parameters are distributed in a population.



Indometacin data for 6 individuals

This individual was one out of six:



To estimate, with the power model, we can choose to

- Estimate individual σ and γ
- Obtain a pooled σ and γ over all individuals.



Pooled estimation of variance parameters

The ML estimate of $(\theta_1, \dots, \theta_n, \sigma, \gamma)$, $n = 6$, is obtained by minimizing

$$\sum_{i=1}^n \left((y_i - f_i(\theta_i))^t (y_i - f_i(\theta_i)) / \sigma^2 f_i(\theta_i)^{2\gamma} + \ln(2\pi\sigma f_i(\theta_i)^\gamma) \right).$$

The generalized nonlinear regression concept is that this is done in alternating steps where θ_i 's are determined, given estimates of γ (σ doesn't matter) by solving

$$f_i'(\theta_i)^t (y_i - f_i(\theta_i)) / f_i(\theta_i)^{2\gamma} = 0, \quad i = 1, \dots, n$$

and given these, updates of γ are obtained from the likelihood above. Iterate until convergence.



How do we summarize the result of 6 individuals? - The Global Two-Stage Method

Our six subjects are only vehicles in our ultimate gain: to get a population summary of the regression parameters.

We therefore assume that the **true** parameter vector for subject i , θ_i is a random sample from a $N(\theta, D)$ distribution, whose parameters we want to estimate.

We only have estimates θ_i^* of θ_i with some precision:

$$\theta_i^* | \theta_i \in N(\theta_i, C_i), \quad \text{so} \quad \theta_i^* \in N(\theta_i, C_i + D)$$

To find θ and D we need a new estimation, by maximizing

$$L_{GTS}(\theta, D) = \sum_{i=1}^N (\ln(\det(C_i + D)) + (\theta_i^* - \theta)^t (C_i + D)^{-1} (\theta_i^* - \theta)).$$



GTS updates regression estimates

Given these we can also obtain refined estimates for θ_i from the equation

$$\theta_i = (C_i^{-1} + D^{-1})^{-1}(C_i^{-1}\theta_i^* + D^{-1}\theta).$$

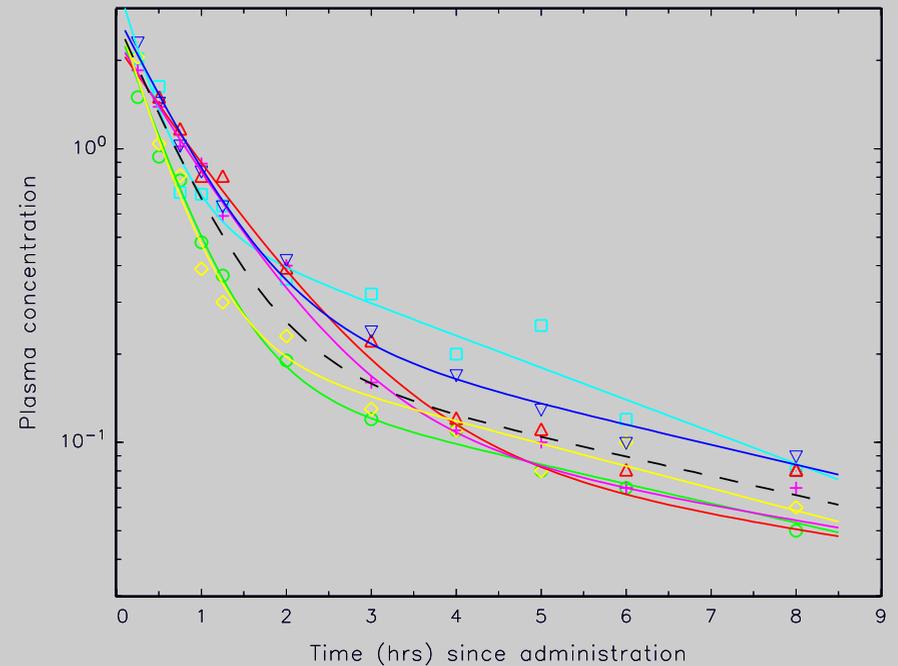
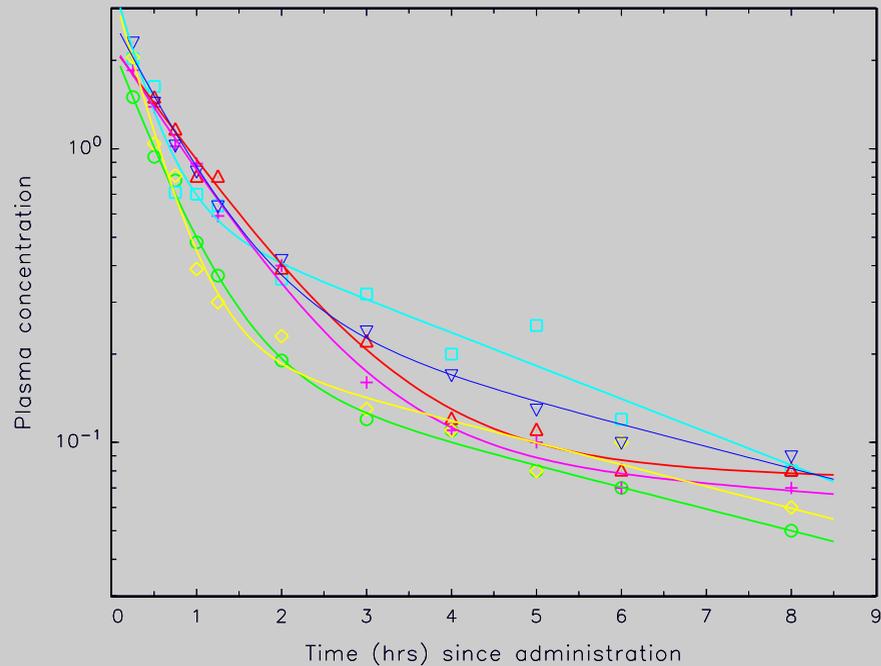
With correct assumptions these could well be better estimates than the original ones.

Table below contains original/updated estimates.

Subj.	$\ln(\theta_1)$	$\ln(\theta_2)$	$\ln(\theta_3)$	$\ln(\theta_4)$
1	0.72/0.92	0.60/0.68	-1.64/-1.72	-1.77/-1.88
2	1.15/1.15	1.05/1.00	-0.40/-0.46	-1.35/-1.38
3	0.77/0.76	-0.04/-0.01	-2.42/-2.19	-4.07/-2.29
4	0.79/0.81	0.07/0.13	-2.31/-2.02	-3.05/-2.19
5	1.22/0.98	0.96/0.82	-1.45/-1.44	-1.76/-1.74
6	0.91/0.95	0.35/0.37	-1.14/-1.23	-1.77/-1.86



Curve fits - original to the left, updated to the right



Note the terminal phases!



GTS summary

The main result is the estimate of θ :

Parameter	Estimate	95% confidence limits	
$\ln(\theta_1)$	0.9273	0.7633,	1.091
$\ln(\theta_2)$	0.4985	0.1755,	0.8216
$\ln(\theta_3)$	-1.511	-2.007,	-1.016
$\ln(\theta_4)$	-1.893	-2.237,	-1.548

(Corresponds to dashed line in last picture - mean parameter curve!)

and

$$D = \begin{pmatrix} 0.016 & 0.044 & 0.073 & 0.039 \\ 0.044 & 0.14 & 0.18 & 0.11 \\ 0.073 & 0.18 & 0.34 & 0.17 \\ 0.039 & 0.11 & 0.17 & 0.09 \end{pmatrix}$$



The Mixed Effects Model setup

Assumptions:

- We have N subjects
- Subject i has n_i observations $y_i = (y_{i1}, \dots, y_{in_i})$
- Assume that $E(y_i | \theta_i) = f_i(\theta_i)$, $V(y_i | \theta_i) = \Lambda_i(\theta_i, \alpha)$ and that the distribution is Gaussian
- Assume that $\theta_i = d_i(\theta, \xi_i)$, $i = 1, \dots, N$, where ξ_i are randomly sampled from a $N(0, D)$ distribution.

In a simple random effects model we take $d_i(\theta, \xi) = \theta + \xi$.



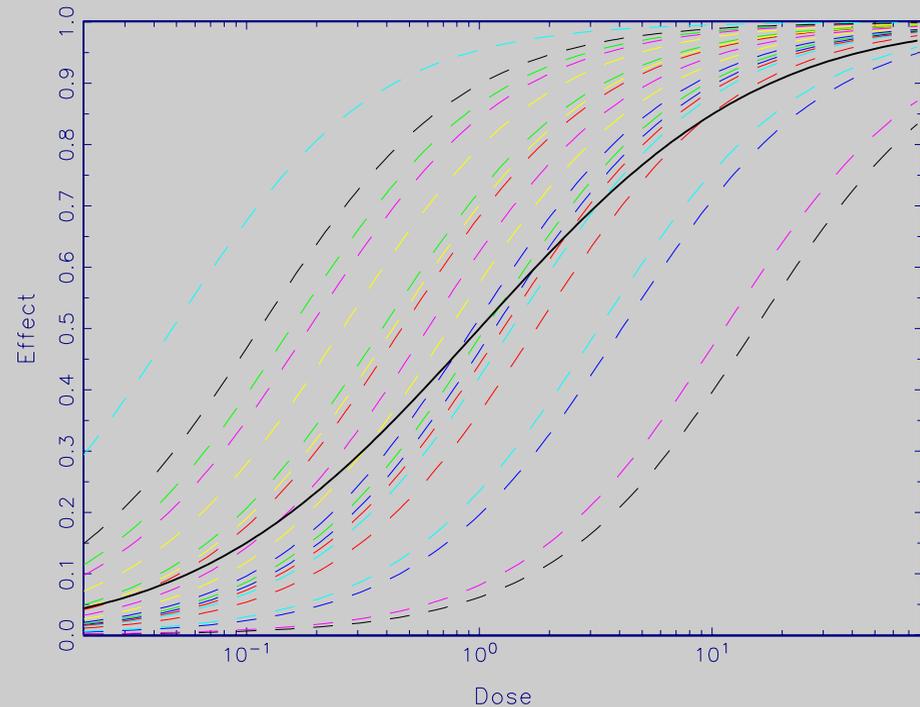
A simulation experiment

We simulated 20 copies of the growth curve function

$$f(d, \xi) = \frac{d}{e^{\xi} + d}$$

where $\xi \in N(0, 2)$. The mean value curve is given approximately by

$$m(d) = \frac{d^{\gamma}}{1 + d^{\gamma}}, \quad \gamma = 0.75.$$



Mean parameter/value curves

It is important to distinguish between the

- Mean Parameter Curve (mpc)

$$y = f(\theta, x) \left(= \frac{d}{1+d} \right)$$

- Mean Value Curve (mvc)

$$y = \frac{1}{\sqrt{\det 2\pi D}} \int f(d(\theta, \xi), x) e^{-\xi^t D^{-1} \xi / 2} d\xi \left(\approx \frac{d^{3/4}}{1+d^{3/4}} \right)$$

as this example shows. Note however that as $D \rightarrow 0$, the mean value curve approaches the mean parameter curve!

Also note that if

$$f(d(\theta, \xi)) = f(\theta) + C(\theta)\xi$$

the two curves coincide!



Soybean experiment

Also the soybean example earlier was part of a bigger experiment:

Data from an experiment to compare the growth patterns of two genotypes of soybean was collected for three consecutive years (1988–1990). Each year 16 plots were planted with seeds, eight with the genotype Forrest (F), a commercial variety, and eight with Plant Introduction #416937 (P), an experimental strain.

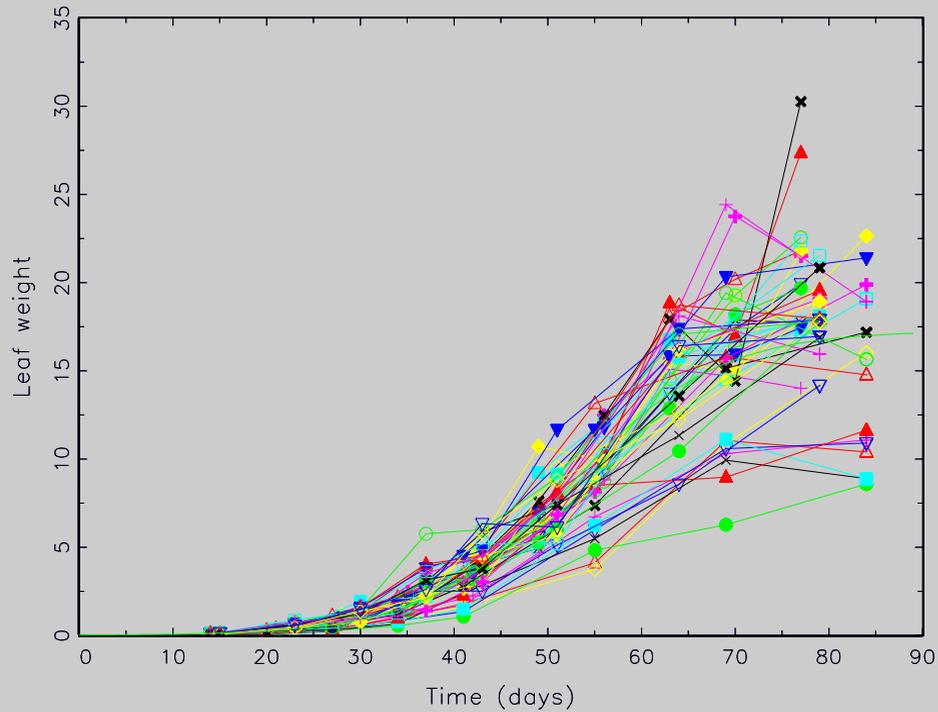
The regression function to fit was

$$y(t) = \frac{\theta_1}{1 + e^{-0.01\theta_3(t-\theta_2)}}.$$

and we use the power model + AR(1)-process for interdependence within plot!



Individual data is shown in

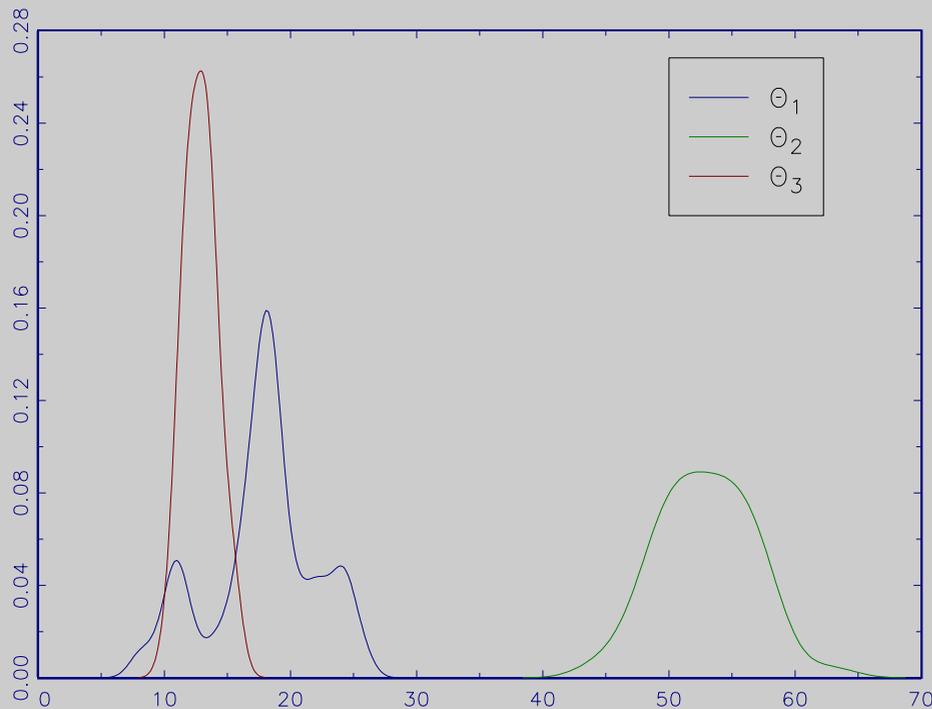


GTS analysis on GLS estimates with pooled variance parameter estimates:

	estimate	standard error	95% confidence limits
θ_1	17.29	0.6138	16.08, 18.49
θ_2	52.56	0.5123	51.56, 53.57
θ_3	12.86	0.1782	12.51, 13.21



Plotting the distribution of the individually estimated coefficients from the GLS analysis....



....so we try to structure data as

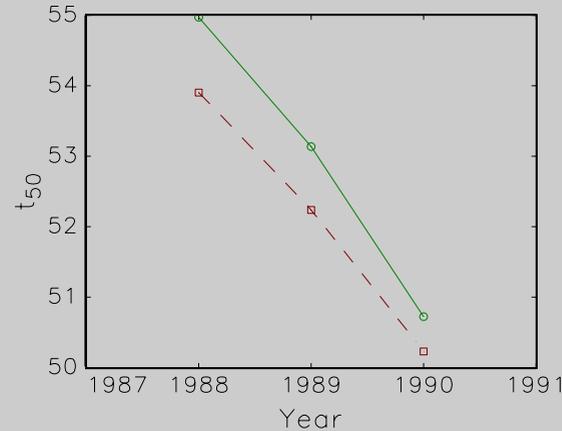
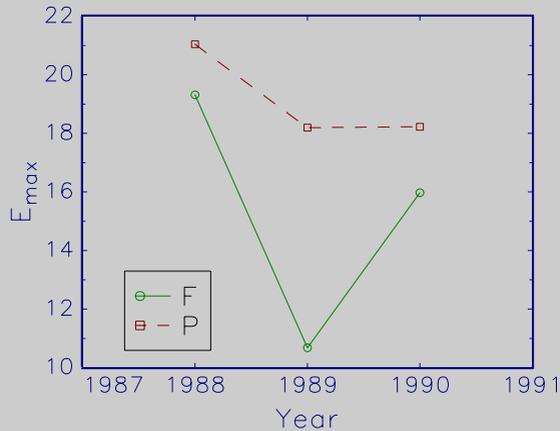
$$\theta_i = A_i\theta + \xi_i, \quad \xi_i \in N(0, D).$$

with one parameter for each genotype \times year combination. (In all $6 \times 3 = 18$ parameters.)

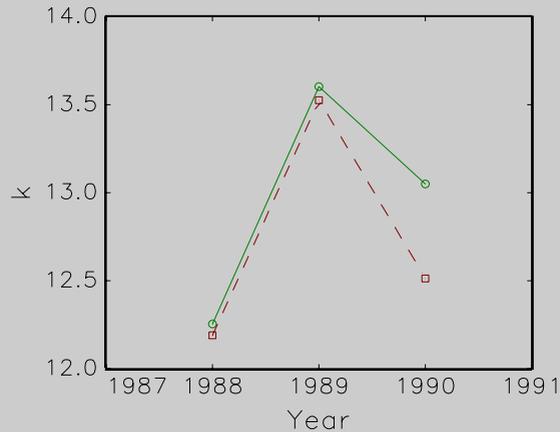


Soybean structured GTS: parameter estimates

The result of the analysis is summarized in the following mean value plot...



...together with



$$D = \begin{pmatrix} 2.198 & 1.952 & -0.5585 \\ 1.952 & 5.018 & -1.661 \\ -0.5585 & -1.661 & 0.5551 \end{pmatrix}$$



Summary of GTS approach to analysis

- Needs rich data - so rich that all individual regression problems can be solved with reasonable values
- Shrinkage of individual coefficients on update – poorly estimated individual coefficients have less weight in the global estimate.
- Some uncertainty about the mathematical logic/model

What we want is a one-step analysis method, so that individual curves need not be well estimated!



The Marginal NONMEM likelihood

The assumptions ...

- Subject i has n_i observations $y_i = (y_{i1}, \dots, y_{in_i})$
- Assume that $E(y_i | \theta_i) = f_i(\theta_i)$, $V(y_i | \theta_i) = \Lambda_i(\theta_i, \alpha)$ and that the distribution is Gaussian

... leads to the following conditional likelihood for subject i :

$$L_i(\theta_i, \alpha | y_i) = e^{-H_i(\theta_i, \alpha, y_i)/2}$$

where

$$H_i(\theta_i, \alpha, y_i) = (y_i - f_i(\theta_i))^t \Lambda(\theta_i, \alpha)^{-1} (y_i - f_i(\theta_i)) + \ln(2\pi \Lambda(\theta_i, \alpha)).$$



The additional assumptions

- $\theta_i = d_i(\theta, \xi_i)$,
- ξ_i random sample of $N(0, D)$ distribution

leads to the marginal likelihood for subject i :

$$L_i^m(\theta, \alpha, D | y_i) = (\det 2\pi D)^{-1} \int L_i(d_i(\theta, \xi), \alpha | y_i) e^{-\xi^t D^{-1} \xi / 2} d\xi$$



The final marginal likelihood becomes

$$L(\theta, \alpha, D) = \prod_{i=1}^N \int e^{-Q_i(\theta, \alpha, D, \xi)/2} d\xi,$$

where

$$Q_i(\theta, \alpha, D, \xi) = (y_i - f_i(d_i(\theta, \xi)))^t \Lambda_i(d_i(\theta, \xi), \alpha)^{-1} (y_i - f_i(d_i(\theta, \xi))) + \ln(2\pi \Lambda(d_i(\theta, \xi), \alpha)) + \xi^t D^{-1} \xi + \ln(\det(2\pi D))$$

This is a multidimensional integral: with 4 regression coefficients, an unstructured D matrix has 6 parameters and we have at least 10 parameters in all.



Lecture 3: Approximative methods

In which we look at different approximations to the marginal likelihood

$$L(\theta, \alpha, D) = \prod_{i=1}^N \int e^{-Q_i(\theta, \alpha, D, \xi)/2} d\xi,$$

where

$$Q_i(\theta, \alpha, D, \xi) = (y_i - f_i(d_i(\theta, \xi)))^t \Lambda_i(d_i(\theta, \xi), \alpha)^{-1} (y_i - f_i(d_i(\theta, \xi))) + \ln(2\pi \Lambda_i(d_i(\theta, \xi), \alpha)) + \xi^t D^{-1} \xi + \ln(\det(2\pi D))$$



Approximating the NONMEM likelihood

Define a new variable $\xi = \xi_i + A_i^{-1/2}\eta$ in integral:

$$\begin{aligned} -2 \ln L_i(\theta, \alpha, D) &= Q_i(\theta, \alpha, D, \xi_i) + \ln(\det A_i) + \\ &-2 \ln \int e^{(Q_i(\theta, \alpha, D, \xi_i) - Q_i(\theta, \alpha, D, \xi_i + A_i^{-1/2}\eta))/2} d\eta. \end{aligned}$$

Want to choose ξ_i and A_i so that the contribution from the integral is minimized. Then we either ignore it or use Gauss-Hermite's integration method to approximate it numerically.

This is done by using that

$$Q_i(\xi_i + A_i^{-1/2}\eta) \approx Q_i(\xi_i) + \partial_\xi Q_i(\xi_i) A_i^{-1/2}\eta + \frac{1}{2}\eta^t A_i^{-t/2} \partial_{\xi\xi}^2 Q_i(\xi_i) A_i^{-1/2}\eta$$

so we want to choose $\xi_i = \xi_i(\theta, \alpha, D)$ to solve

$$\partial_\xi Q_i(\theta, \alpha, D, \xi) = 0, \quad \text{and choose} \quad A_i = -E(\partial_{\xi\xi}^2 Q_i(\theta, \alpha, D, \xi_i))/2$$



Introduce notations

$$\mu_i(\theta, \xi) = f_i(d_i(\theta, \xi)), \quad C_i(\theta, \xi) = \partial_\xi \mu_i(\theta, \xi).$$

and assume that Λ_i does not depend on ξ . (From now on we will cheat by replacing $\Lambda_i(d_i(\theta, \xi_i), \alpha)$ with $\Lambda_i(\theta, \alpha) = \Lambda_i(d_i(\theta, 0), \alpha)$.)

Then $\xi_i = \xi_i(\theta, \alpha, D)$ is given as the solution of

$$-0.5 \partial_\xi Q_i(\theta, \alpha, D, \xi) = C_i(\theta, \xi)^t \Lambda_i^{-1}(y_i - \mu_i(\theta, \xi)) - D^{-1} \xi = 0$$

and

$$A_i = -E(\partial_{\xi\xi}^2 Q_i(\theta, \alpha, D, \xi_i))/2 = C_i(\theta, \xi_i)^t \Lambda_i^{-1} C_i(\theta, \xi_i) + D^{-1}.$$

We can choose these also if Λ_i depend on ξ – though at the expense of less precision.

Then we have that

$$\int e^{(Q_i(\theta, \alpha, D, \xi_i) - Q_i(\theta, \alpha, D, \xi_i + A_i^{-1/2} \eta))/2} d\eta = \int e^{R_i(\theta, \alpha, D, \xi_i, \eta)} e^{-|\eta|^2/2} d\eta,$$

where R_i is approximately zero to second order in $\eta = 0$.



Note that we eliminate the integral if $Q(\xi)$ is a second order polynomial (and Λ_i does not depend on ξ). This is the case if

$$\mu_i(\theta, \xi) = \mu_i(\theta) + C_i(\theta)\xi,$$

and Λ_i does not depend on ξ , which is called the *quasi-linear case* and contains the linear case as a special case.



The Laplacian approximation

The non-integral part of the likelihood can now be written^a

$$-2 \ln L(\theta, \alpha, D) = \sum_{i=1}^N \nu_i(\theta, \alpha, D)^t V_i(\theta, \alpha, D)^{-1} \nu_i(\theta, \alpha, D) + \ln(\det V_i(\theta, \alpha, D)), \quad (3)$$

where

$$\nu_i(\theta, \alpha, D) = y_i - \mu_i(\theta, \xi_i) + C_i(\theta, \xi_i) \xi_i,$$

and

$$V_i(\theta, \alpha, D) = \Lambda_i(\theta, \alpha) + C_i(\theta, \xi_i) D C_i(\theta, \xi_i)^t$$

This is the *Laplacian* approximation, which requires the computation of ξ_i each time the objective function is computed - big computational burden.

^aleft as an exercise – not trivial, requires some matrix operations



Conditional First Order approximation

is obtained by taking subject specific choices for ξ_i , holding them fixed and minimizing (3) with respect to θ, α and D .

This means iterating between

1. given (θ, α, D) , estimate $\xi_i, i = 1, \dots, N$ by minimizing

$$PLS(\xi_1, \dots, \xi_N) = \sum_{i=1}^N Q_i(\theta, \alpha, D, \xi_i) \quad (\text{Penalized least squares})$$

2. minimize (3) given these individual estimates.

until convergence in (θ, α, D) .



Population average or First Order approximation

is obtained by taking $\xi_i = 0$ in (3). This leads to the likelihood

$$-2 \ln L(\theta, \alpha, D) = \sum_{i=1}^N (y_i - \mu_i(\theta))^t V_i^{-1} (y_i - \mu_i(\theta)) + \ln(\det V_i),$$

where

$$V_i = \Lambda_i + C_i(\theta) D C_i(\theta)^t$$

(when ξ is dropped as an argument, it means computed at $\xi = 0$).

Again this is the likelihood in the quasi-linear case! It is called the first order (FO) approximation to the likelihood because it is the likelihood when we approximate $\mu_i(\theta, \xi)$ with its linear part

$$\mu_i(\theta) + C_i(\theta)\xi$$



Comparison between ML, Laplace and FO methods

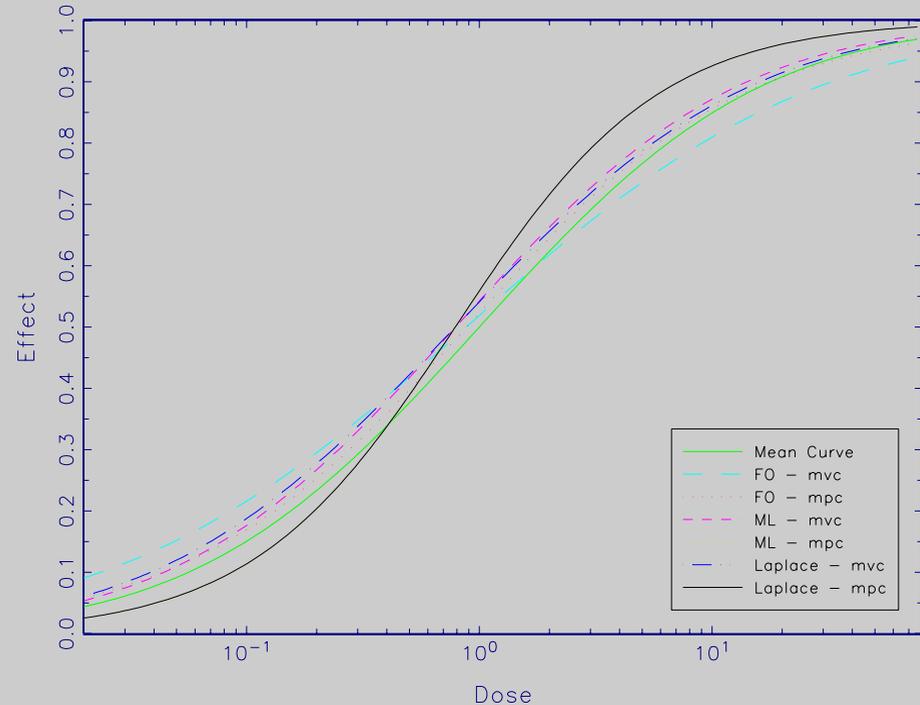
Estimate model $E = \frac{d^\gamma}{e^\theta + d^\gamma}$, with $\theta = \theta_0 + \xi$ random:

True values: $\gamma = 1, \theta_0 = 0, D = 2$

Sample of 20 independent observations

Parameter estimates:

Method	θ_0	γ	D	σ
ML	-0.24	1.00	2.02	0.11
Laplace	-0.24	0.99	2.43	0.11
FO	-0.13	0.73	2.33	0.12



- Mean value curve is not the same as mean parameter curve!
- FO approximates regression function to mean data!

In fact, the FO method fits the quasilinear model

$$f(d | \theta, \gamma, \xi) = \frac{d^\gamma}{e^\theta + d^\gamma} - \frac{d^\gamma e^\theta}{(e^\theta + d^\gamma)^2} \xi$$

which is a good approximation to

$$\frac{d^\gamma}{e^{\theta+\xi} + d^\gamma}$$

only when $\xi \approx 0$ (i.e. covariance matrix D small).

The more difference between the mean parameter curve and the mean value curve, the more in doubt should results from the first order approximation be, except as a description of the population mean behavior.



Example: Indomethacin data

Here are the estimates of the regression parameters for a few different methods:

GTS

Parameter	Estimate	95% confidence limits	
$\ln(\theta_1)$	0.9273	0.7633,	1.091
$\ln(\theta_2)$	0.4985	0.1755,	0.8216
$\ln(\theta_3)$	-1.511	-2.007,	-1.016
$\ln(\theta_4)$	-1.893	-2.237,	-1.548

Conditional First Order

Parameter	Estimate	95% confidence limits	
$\ln(\theta_1)$	0.9066	0.7896,	1.024
$\ln(\theta_2)$	0.5492	0.423,	0.6753
$\ln(\theta_3)$	-1.335	-1.669,	-0.9996
$\ln(\theta_4)$	-1.716	-1.978,	-1.453

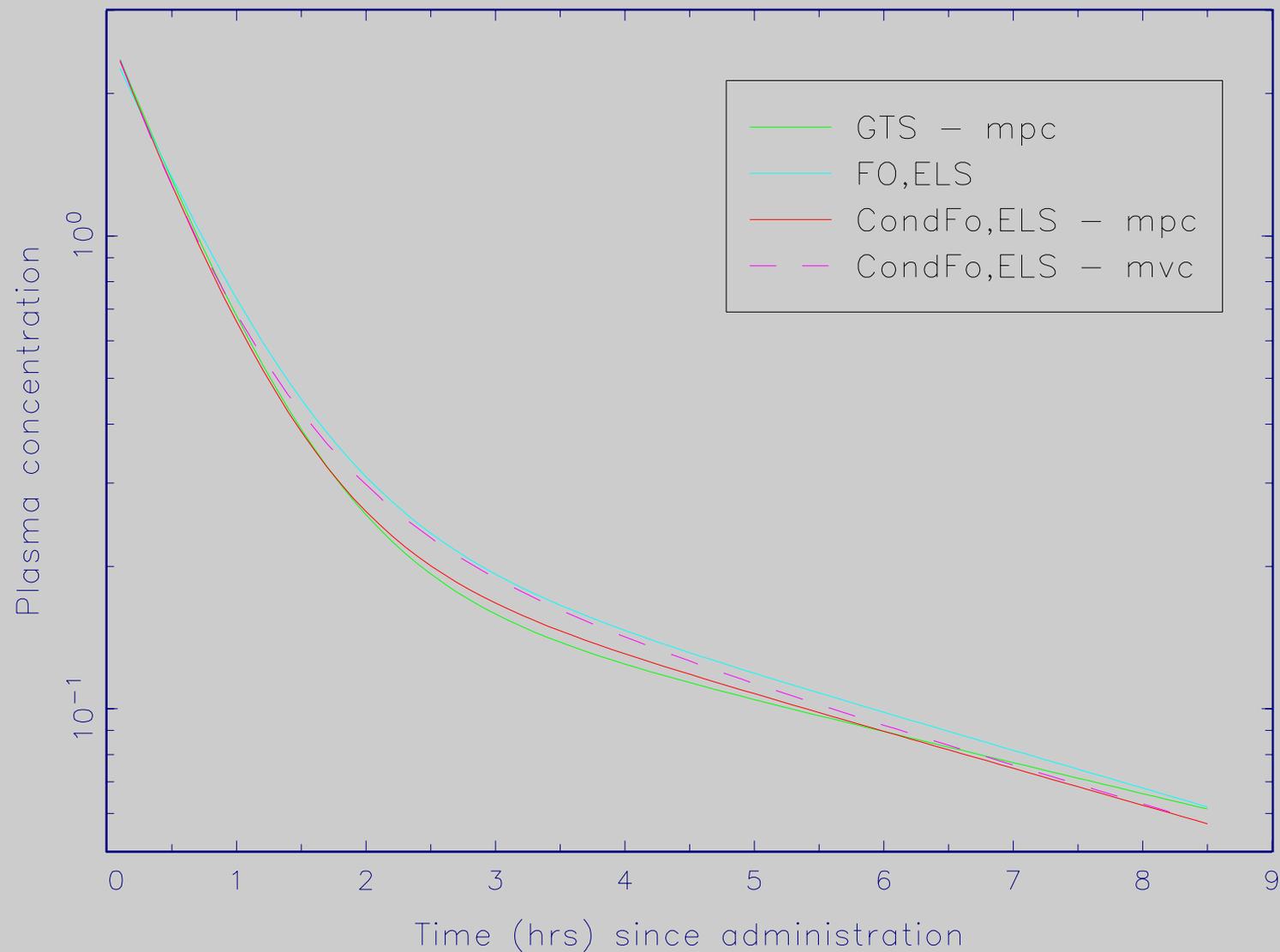
First Order

Parameter	Estimate	95% confidence limits	
$\ln(\theta_1)$	0.8318	0.7137,	0.95
$\ln(\theta_2)$	0.4335	0.2106,	0.6564
$\ln(\theta_3)$	-1.22	-1.689,	-0.752
$\ln(\theta_4)$	-1.695	-2.002,	-1.388

Rather similar results!



A graphical display of the difference can be shown as mean parameter curves:



Population covariance matrix (correlation below diagonal) and other variability measurements:

GTS

$$D = \begin{pmatrix} 0.016 & 0.044 & 0.073 & 0.039 \\ 0.92 & 0.14 & 0.18 & 0.11 \\ 0.98 & 0.81 & 0.34 & 0.17 \\ 0.997 & 0.94 & 0.96 & 0.09 \end{pmatrix}$$

$$\sigma = 0.130, \gamma = 0.954$$

Conditional First Order

$$D = \begin{pmatrix} 0.014 & 0.035 & 0.048 & 0.021 \\ 0.88 & 0.11 & 0.093 & 0.056 \\ 0.93 & 0.65 & 0.19 & 0.068 \\ 0.99 & 0.94 & 0.88 & 0.033 \end{pmatrix}$$

$$\sigma = 0.1234, \gamma = 0.7691$$

First Order

$$D = \begin{pmatrix} 0.0092 & 0.033 & 0.055 & 0.036 \\ 0.96 & 0.13 & 0.17 & 0.12 \\ 0.88 & 0.72 & 0.42 & 0.25 \\ 0.96 & 0.85 & 0.98 & 0.15 \end{pmatrix}$$

$$\sigma = 0.1192, \gamma = 0.8568$$

surprisingly similar results,

which must be due to 'small' D



From FO NONMEM to FO GEE

The FO NONMEM case is the likelihood

$$-2 \ln L(\theta, \alpha, D) = \sum_{i=1}^N (y_i - g_i(\theta))^t V_i^{-1} (y_i - g_i(\theta)) + \ln(\det V_i),$$

$$V_i = \Lambda_i + C_i(\theta) D C_i(\theta)^t.$$

Let $\omega = (D, \alpha)$. Then the estimating equations are:

$$\begin{cases} \sum_{i=1}^N g_i'(\theta)^t V_i^{-1} (y_i - g_i(\theta)) + \frac{1}{2} \left(\frac{\partial v_i}{\partial \theta} \right)^t (V_i^{-1} \otimes V_i^{-1}) (s_i(\theta) - v_i(\theta, \omega)) & = 0 \\ \frac{1}{2} \sum_{i=1}^N \left(\frac{\partial v_i}{\partial \omega} \right)^t (V_i^{-1} \otimes V_i^{-1}) (s_i(\theta) - v_i(\theta, \omega)) & = 0, \end{cases}$$

where

$$s_i(\theta) = \text{vec}((y_i - g_i(\theta))(y_i - g_i(\theta))^t), \quad v_i(\theta, \omega) = \text{vec}(V_i(\theta, \omega)).$$

The first equation is a compromise between first **two** statistical moments.



In the GEE approach we focus on the regression part:

$$\begin{cases} \sum_{i=1}^N g'_i(\theta)^t V_i^{-1} (y_i - g_i(\theta)) & = 0 \\ \frac{1}{2} \sum_{i=1}^N \left(\frac{\partial v_i}{\partial \omega}\right)^t (V_i^{-1} \otimes V_i^{-1}) (s_i(\theta) - v_i(\theta, \omega)) & = 0. \end{cases}$$

Solved by iterating (given θ , estimate ω from 2nd eqn and then update θ from first).

Is actually a Fixed Point problem: Define the function $\Psi(\theta)$ as follows: solve for ω and use this to update to a new estimate of θ , which then is $\Psi(\theta)$.



Theophylline data

Administered orally to twelve subjects, and serum concentrations were measured at ten time points per subject over the subsequent 24 hours.

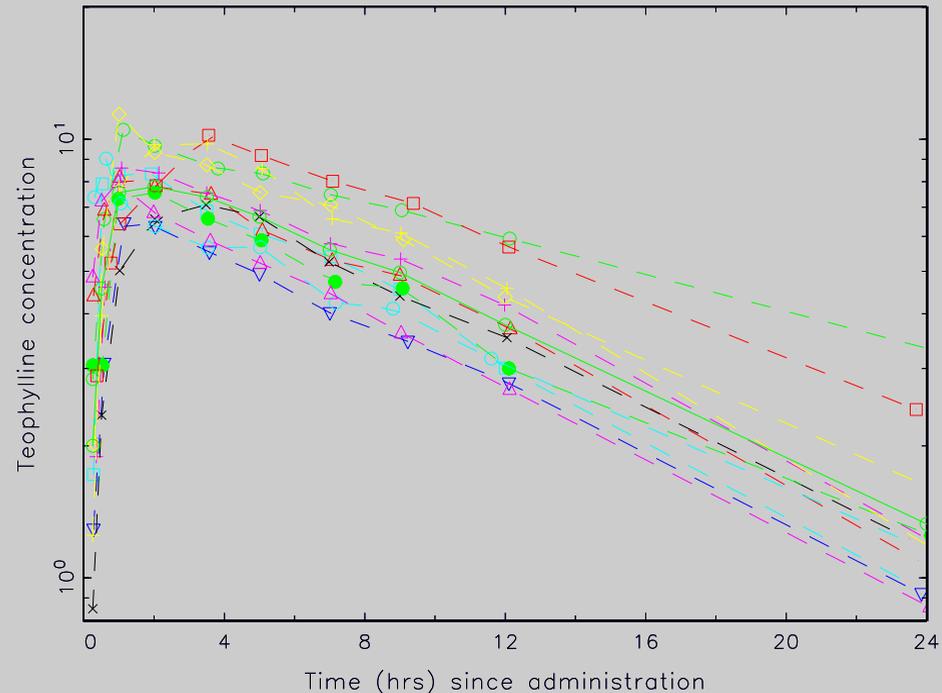
The model for subject i is that $C_i(t)$ is given by

$$\frac{D_i k_{ai}}{V_i(k_{ai} - CL_i/V_i)} (e^{-CL_i t/V_i} - e^{-k_{ai} t}).$$

All three regression parameters are assumed random.

This model corresponds to 1-compartment model with first order absorption:

$$V_i C_i'(t) = k_{ai} D_i e^{-k_{ai} t} - CL_i C_i(t).$$

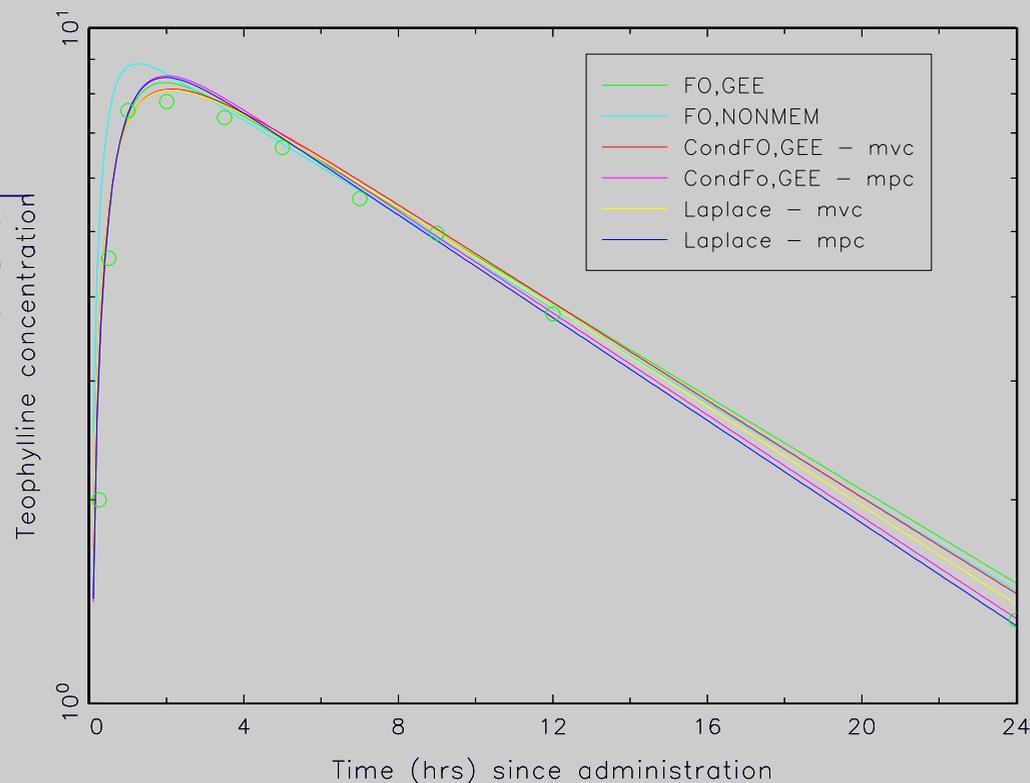


NONMEMs on Theophylline data

We estimated this model (with the Power model) under a variety of different models:

Model	k_a	Cl	V	σ	γ
FO, GEE	1.62	0.038	0.476	0.4345	0.33
FO, NONMEM	2.97	0.038	0.472	0.098	1.00
CondFO, GEE	1.51	0.040	0.457	0.3554	0.43
Laplace	1.54	0.040	0.460	0.4047	0.35

- Not much difference between Mean Parameter Curve and Mean Value Curve
- We see the NONMEM-effect: poorer fit around C_{max} - but that part is poorly represented in data
- The GEE approach is safer, simpler and faster than the NONMEM approach.



CondFo, GEE is similar to NLME

The NLME method, by Bates and Lindstrom, replaces the variance parameter estimation, which, for the GEE method, is done by minimization of

$$-2 \ln L(\alpha, D) = \sum_{i=1}^N (\nu_i^t V_i^{-1} \nu_i + \ln(\det V_i)),$$

where

$$\nu_i = y_i - \mu_i(\theta_p, \xi_i) + C_i(\theta_p, \xi_i)\xi_i,$$

by the corresponding linear mixed effects model, which minimizes

$$-2 \ln L(\delta\theta, \alpha, D) = \sum_{i=1}^N (\nu_i - A_i(\theta_p, \xi_i)\delta\theta)^t V_i^{-1} (\nu_i - A_i(\theta_p, \xi_i)\delta\theta) + \ln(\det V_i),$$

Here θ_p is the previous value of θ , $\delta\theta = \theta - \theta_p$ and $A_i(\theta, \xi) = \partial_{\theta}\mu_i(\theta, \xi)$.



Vonesh-Carters approach

Here we replace the parametric variance estimation of the GEE estimate with a simple, non-parametric estimate. Requires rich data to work!

To estimate an individual ξ_i , we solve

$$C_i(\theta, \xi)^t \Lambda_i^{-1} (y_i - \mu_i(\theta, \xi)) = 0.$$

Can be obtained numerically by a Fisher scoring equation:

$$\xi = \xi + (C_i(\theta, \xi)^t \Lambda_i^{-1} C_i(\theta, \xi))^{-1} C_i(\theta, \xi)^t \Lambda_i^{-1} (y_i - \mu_i(\theta, \xi)).$$

Vonesh-Carters approach takes one step and starts at $\xi = 0$.

Based on these individual estimates, we obtain a first estimate of D : let

- $\hat{\xi}_i$ be the individual estimate
- ξ_i be the true individual parameter, with $V(\xi_i) = D$.
- Assume that ξ_i and $\hat{\xi}_i - \xi_i$ are independent.



An estimate of the variance of $\hat{\xi}_i - \xi_i$ is

$$V(\hat{\xi}_i - \xi_i) = \frac{1}{n} \sum_{i=1}^n (C_i^t \Lambda_i^{-1} C_i)^{-1}.$$

From this we derive an estimate for D of the form

$$D^* = V(\hat{\xi}_i) - \frac{t}{n} \sum_{i=1}^n (C_i^t \Lambda_i^{-1} C_i)^{-1},$$

for a number t that assures the positive definiteness of D^* .



Vonesh-Carters method on Theophylline data

Comparison of Vonesh-Carter and parametric GEE:

Model	k_a	Cl	V	σ	γ
FO, GEE	1.62	0.038	0.476	0.4345	0.33
FO, VC	1.57	0.038	0.473	0.2943	0.32
CondFO, GEE	1.51	0.040	0.457	0.3554	0.43
CondFO, VC	1.50	0.039	0.461	0.2897	0.37

In this case - not much difference, except on the variability side!



Lecture 4: Population pharmacokinetics

In which we look at the original motivator for NONMEMs: population pharmacokinetics and also discuss the robust estimator of the covariance of parameter estimates.



- Pharmacokinetics (PK) is the description of how the body handles a drug.
- Historically this was originally done in terms of different compartmental models, leading to systems of differential equations
- During the 70's Malcolm Rowland and others transformed the focus to the estimation of key parameters: clearance and volume of distribution.
- To describe the PK of a drug, we usually have rather rich data in a small number of subjects (see indomethacin and theophylline data discussed previously), but
- in order to describe variations of PK in a larger population, odd blood sampling is taken in larger clinical trials.



Population pharmacokinetics

“Design:” blood sampling, at random both in time and number of assessments, in patients treated with a particular drug.

Leads to sparse data, which were very hard to do anything with at all. Motivated Lou Sheiner and Stuart Beals to develop the package NONMEM. Very much used by pharmacokineticists.

Modelling consists of a (often very simple) model for the pharmacokinetics, and then various parameters – notably the clearance – is investigated for its dependence on various covariates (like age, sex and weight).



Example: Phenobarbital data

Data (155 samples, 1–6/subject) were collected in 59 pre-term infants given phenobarbital for prevention of seizures during the first 16 days after birth.

<i>Subject 9</i>			<i>Subject 50</i>		
<i>time</i>	<i>dose</i>	<i>conc.</i>	<i>time</i>	<i>dose</i>	<i>conc.</i>
(hrs)	($\mu\text{g /kg}$)	($\mu\text{g /L}$)	(hrs)	($\mu\text{g /kg}$)	($\mu\text{g /L}$)
0.0	27.0	.	0.0	20.0	.
1.1	.	22.1	3	.	22.2
11.1	3.2	.	12.5	2.5	.
22.3	3.2	.	24.5	2.5	.
34.6	3.2	.	36.5	2.5	.
⋮	⋮	.	⋮	⋮	.
82.7	.	29.2	81.0	.	30.5
83.2	3.2	.	84.5	2.5	.
94.6	3.2	.	88.0	30.0	.
142.6	3.2	.	132.5	3.5	.
312.6	.	19.6	144.5	3.5	.
			157.0	3.5	.
			162.0	.	58.7
Apgar	8		Apgar	6	
Weight	1.4 kg		Weight	1.1 kg	



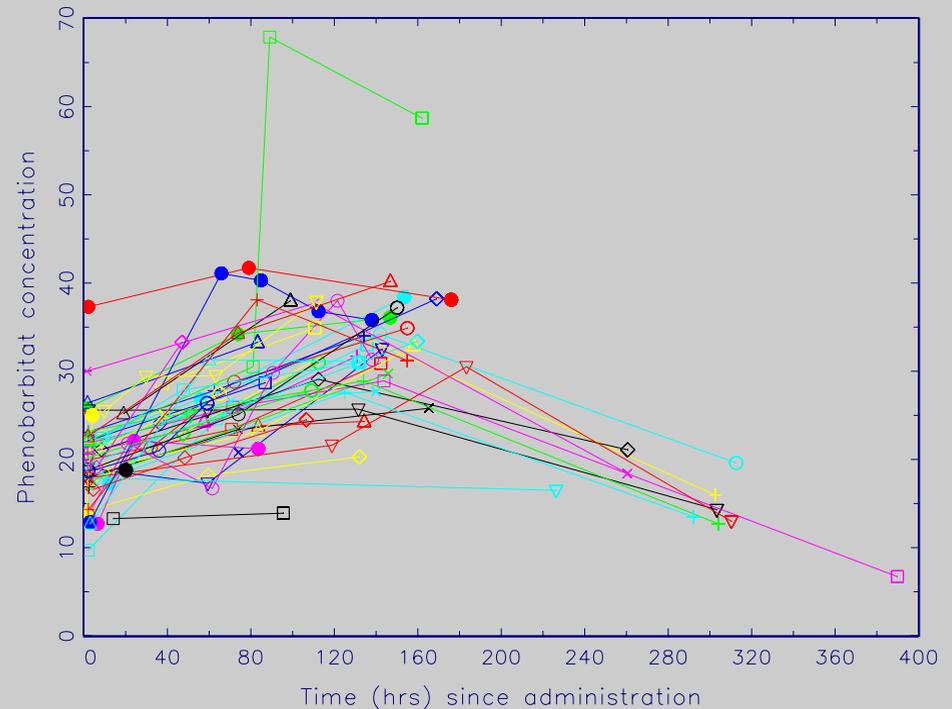
We model the drug in the body with a one-compartment PK-model, which has two intrinsic parameters: clearance (CL) and volume (V):

$$VC'(t) = \alpha(t) - CL \cdot C(t)$$

where $\alpha(t)$ is the amount absorbed per time unit.

If CL_i and V_i are clearance and distribution volume for subject i , and single i.v. doses D_{id} were administered to subject i at times t_{id} , we have

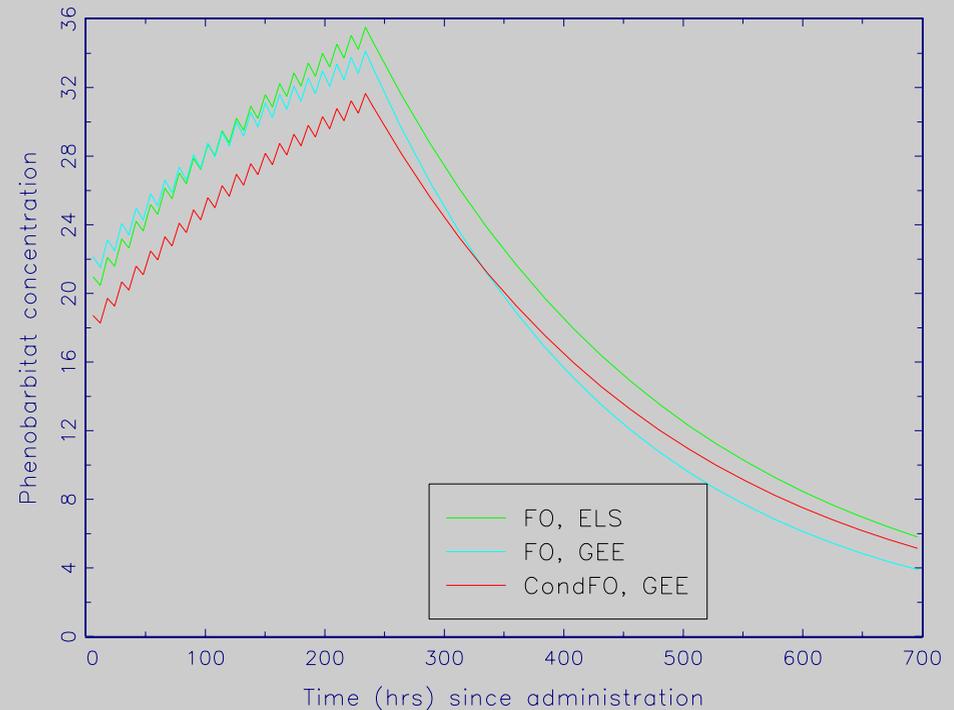
$$C(t) = \sum_{d:t_{id}<t} \frac{D_{id}}{V_i} \exp(-CL_i(t - t_{id})/V_i).$$



For the NONMEM model we start with a simple model, assuming diagonal D , independent, identical errors within subject and that all regression coefficients are random.

Result:

Parameter	FO	FO	CondFO
	NONMEM	GEE	GEE
CL_i	5.482	6.190	6.159
V_i	1.398	1.318	1.566
σ	2.83	2.80	3.20
D_{11}	6.845	8.269	5.314
D_{22}	0.2857	0.2339	0.7511

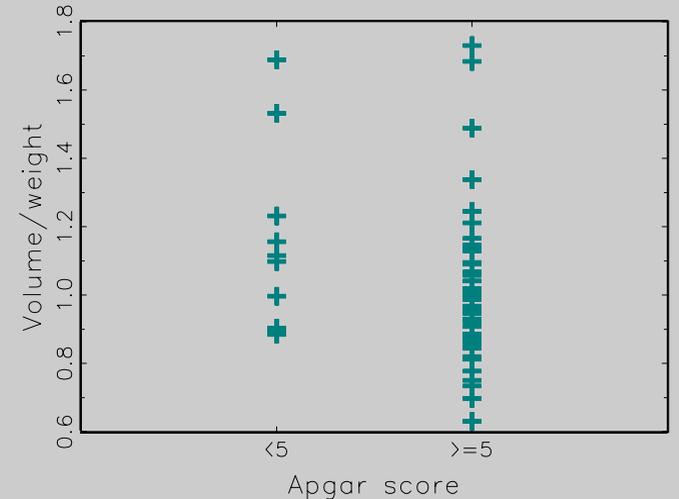
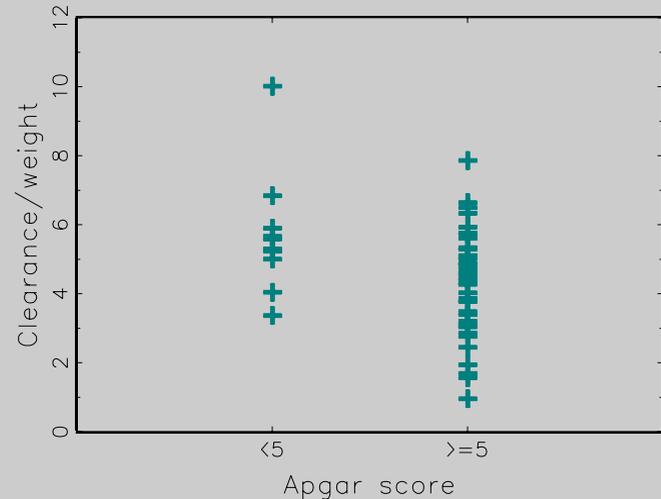
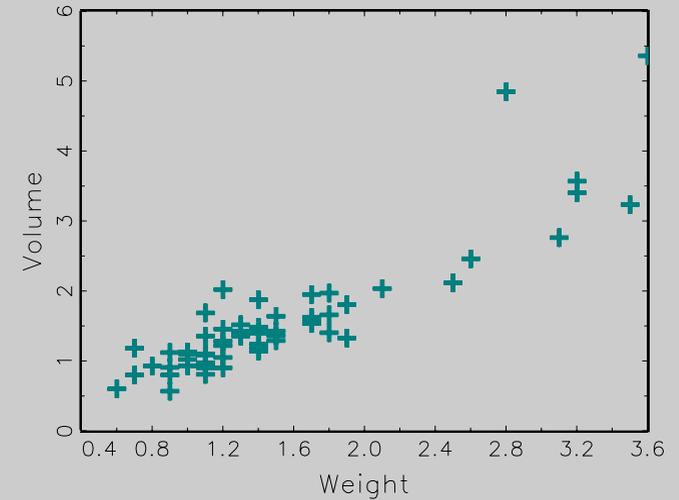
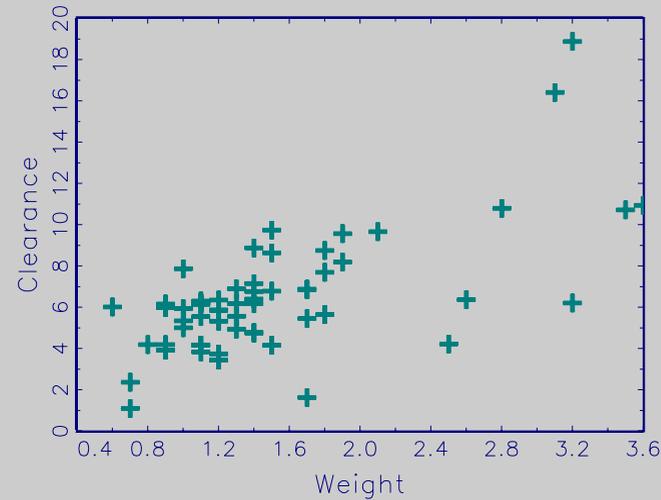


Predictions from loading dose of 30 mg, followed by 3 mg every 12th hour.



We can plot the *individual estimates* obtained, vs the covariates weight and Apgar score.

Both CL and V appear proportional to baby weight and there is possibly an additional effect of a low Apgar score on V .



This leads to the following model, which uses that both parameters are positive:

$$CL_i = \theta_1 w_i e^{\xi_{1i}}, \quad V_i = \theta_2 w_i (1 + \theta_3 I_{[\text{Apgar} < 5]}) e^{\xi_{2i}},$$

(w_i is the weight of the baby) together with a multiplicative Power model for within-baby error structure.

Parameter	FO, NONMEM	FO, GEE	CondFO, NONMEM	CondFO, GEE	ML
θ_1	4.676	4.656	4.656	4.637	4.617
θ_2	0.9649	0.9670	1.003	0.9682	0.9752
θ_3	0.1524	0.1478	0.1564	0.1520	0.1587
σ	0.1038	0.1038	0.1086	0.1068	0.1151
D_{11}	0.04318	0.4348	0.3822	0.03688	0.02897
D_{22}	0.02706	0.02709	0.03001	0.02784	0.02787



The robust estimator of covariance

If we estimate parameters in a statistical model using a likelihood, we often use the negative of the inverse of the hessian (or its expected value, the Fisher information $I(\theta)$) as the covariance matrix for the estimates.

However using that, assumes that the model is correct! Another approach consists of

- Justify a estimating equation that should produce sensible parameter estimates, but need not correspond to a correct probability model
- Use a covariance estimator for the parameter estimates that is (essentially) correct even if the model is wrong.

This approach is the so-called GEE approach to clustered data (GEE = Generalized Estimating Equations) by Liang and Zeger. Is used also when analyzing multiple occurrences with survival models (Cox Regression).



Assume that we estimate a parameter from an equation

$$G(\theta) = 0.$$

(Can be, but need not be, the score equation of a likelihood.) If the solution is $\hat{\theta}$, how can we obtain estimates for $V(\hat{\theta})$?

Let θ be the “true” parameter value. First order approximation around $\hat{\theta}$ then gives us that

$$G(\theta) \approx G(\hat{\theta}) + G'(\hat{\theta})(\theta - \hat{\theta}) = G'(\hat{\theta})(\theta - \hat{\theta})$$

from which we can deduce that

$$V(G(\theta)) \approx E(G'(\hat{\theta}))V(\hat{\theta})E(G'(\hat{\theta}))^t$$

This leads to the sandwich or robust estimator.

$$V(\hat{\theta}) \approx E(G'(\hat{\theta}))^{-1}V(G(\theta))E(G'(\hat{\theta})^t)^{-1}$$



This can be used in the estimation of confidence intervals etc for the estimate θ , and is a partial guard against unwanted effects of model miss-specification when $V_i(\theta)$ is replaced by W_i .

For
$$G(\theta) = \sum_{i=1}^N \mu_i'(\theta)^t W_i^{-1} (y_i - \mu_i(\theta)),$$

the variance is given by
$$V(G(\theta)) = \sum_{i=1}^N \mu_i'(\theta)^t W_i^{-1} V(y_i) W_i^{-1} \mu_i'(\theta).$$

Here $V(y_i)$ can be estimated by
$$(y_i - \mu_i(\hat{\theta}))(y_i - \mu_i(\hat{\theta}))^t,$$

but there are better estimates (more unbiased).

Moreover,
$$E(G'(\theta)) = - \sum_{i=1}^N \mu_i'(\theta)^t W_i^{-1} \mu_i'(\theta).$$

The robust estimator is now easy to compute!



Example: Theophylline data

Instead of complicated NONMEMs, could we use OLS as estimating equation and use the robust estimator? This analysis on the theophylline data 4.8 shows

OLS with robust estimator

	estimate	standard error	95% confidence limits
$\ln k_a$	0.3982	0.1716	0.06184, 0.7345
$\ln CL$	-3.248	0.07641	-3.398, -3.098
$\ln V$	-0.7237	0.04068	-0.8035, -0.644

NONMEM, Vonesh-Carters method

	estimate	standard error	95% confidence limits
$\ln k_a$	0.4615	0.221	0.02836, 0.8946
$\ln CL$	-3.265	0.0896	-3.441, -3.09
$\ln V$	-0.7477	0.03879	-0.8237, -0.6716

Not much difference?



Lecture 5: Some examples from Clinical Trials

In which we look at three examples where the lecturer have used/experimented with nonlinear mixed effects models in non-PK clinical data.

1. Six hour ACTH stimulation to assess cortical dysfunction after steroid treatment
2. Analysis of diary card data (peak flow), to find time to onset of action
3. Dose response in the presence of a partial agonist – do salmeterol compromise the response to an inhaled short-acting β_2 -agonist?



A six hour ACTH infusion test of the adrenal glands

Background: Adrenal stimulation is often used to assess HPA-axis function, and therefore to determine the evaluate the effect of glucocorticosteroids (GCSs) on the adrenal glands. The ability of the adrenal glands to produce and secrete cortisol is measured by plasma cortisol levels.

Objective: To related the effect of Pulmicort Turbuhaler (b.i.d.) on the HPA-axis to that of oral prednisolone treatment (given once daily), using plasma cortisol levels during a six hour ACTH stimulation.

Study Design: Parallel group, one week screening, one week baseline period followed by six weeks of treatment. ACTH test at randomization and after six weeks. The study was double blind and double-dummy. Cortisol measurements taken before and 2,4 and 6 hours after initiation of ACTH stimulation.

Investigational Treatments: Pulmicort Turbuhaler in daily doses 800, 1600 and 3200 μg , prednisolone 10 mg daily dose and placebo.



Model Equation Assuming that cortisol follows a one-compartmental model we start from the equation

$$C'(t) = \alpha(t) - kC(t),$$

where $\alpha(t)$ is the production rate (amount per hour and liter). Normally $\alpha(t)$ has a diurnal rhythm, peaking in the morning and being low during most of the early night.

During ACTH stimulation we assume $\alpha(t) = I$.

Can solve equation to

$$C(t) = C_0e^{-kt} + \frac{I}{k}(1 - e^{-kt})$$

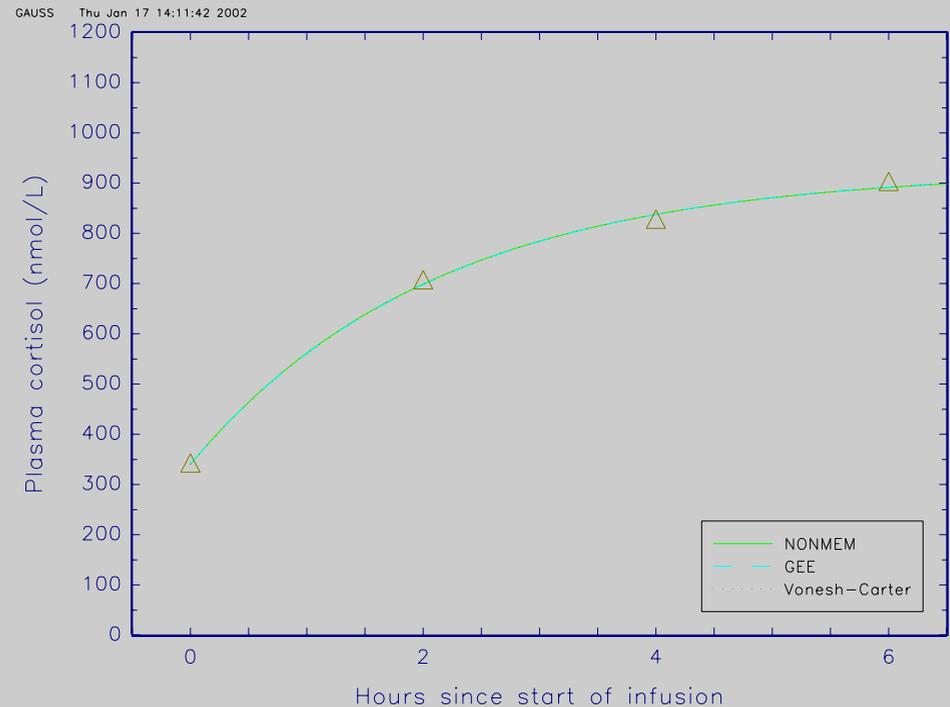
Three parameters: (C_0, I, k) . We will assume k is a population parameter, so that the problem becomes quasi-linear!



Analysis of baseline data

Parameter estimates using three first order methods:

parameter	NONMEM	GEE	Vonesh-Carter
I	439.5	434.9	435.0
k	0.4751	0.4690	0.4690
C_0	340.2	340.6	340.6
D_{11}	2550	2503	2551
D_{12}	537.8	521.9	527.2
D_{22}	8748	8731	8926
σ	46.94	46.96	46.97



No difference in estimates and consequently not in the mean parameter curves = mean value curves. (Rich and well-behaved data!)



Note that, if the model is correct, NONMEM produces likelihood estimates in this case. Using information matrix for standard errors.

NONMEM method:

	estimate	standard error	95% confidence limits
l	439.5	19.59	401.1, 477.9
k	0.4751	0.02428	0.4275, 0.5227
C ₀	340.2	14.23	312.3, 368.1

GEE method:

	estimate	standard error	95% confidence limits
l	434.9	17.91	399.8, 470
k	0.469	0.02219	0.4255, 0.5125
C ₀	340.6	14.68	311.8, 369.4

Vonesh-Carters method:

	estimate	standard error	95% confidence limits
l	435.0	17.91	399.9, 470
k	0.469	0.02219	0.4255, 0.5125
C ₀	340.6	14.68	311.8, 369.4

Almost identical results!



Analysis of six-week ACTH experiment

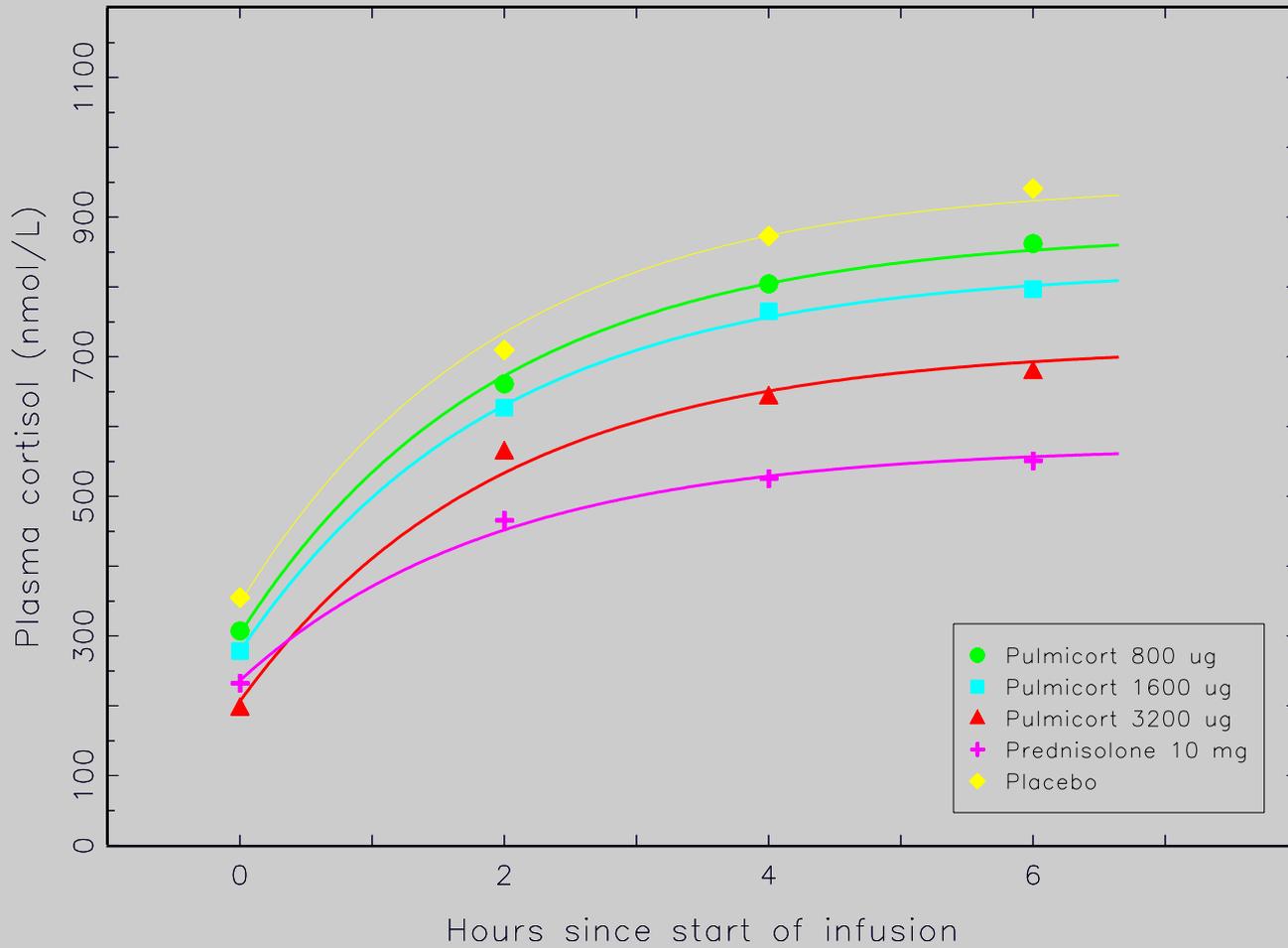
Design model: C_0 and I are assumed to be treatment specific: (C_0, I) for a patient in treatment group i takes the form $m_i + \xi$, where $\xi \in N(0, D)$, where m_i is a 2-vector and D is a 2×2 covariance matrix!

In all $5 \times 2 + 1 = 11$ mean value parameters!

NONMEM method, since it is the correct likelihood method in this case.



Graphical display of result



Analysis results of six-week ACTH experiment

Done both with the hessian (left)

and the robust (right) covariance estimator:

Parameter	estimate	standard error	95% confidence limits	standard error	95% confidence limits
I_{800}	450.8	26.84	398.2, 503.4	34.23	383.7, 517.9
I_{1600}	424.3	26.76	371.8, 476.7	33.1	359.4, 489.1
I_{3200}	367.4	25.13	318.1, 416.6	29.21	310.1, 424.6
I_{pred}	293.5	22.26	249.9, 337.1	19.32	255.6, 331.3
I_{pbo}	487.8	28.03	432.9, 542.8	24.85	439.1, 536.6
k	0.5129	0.02597	0.462, 0.5638	0.02989	0.4543, 0.5715
$C_{0,800}$	303.2	31.07	242.3, 364.1	24.18	255.8, 350.5
$C_{0,1600}$	277.9	32.58	214.1, 341.8	49.22	181.5, 374.4
$C_{0,3200}$	206.4	32.57	142.5, 270.2	41.97	124.1, 288.6
$C_{0,pred}$	236.2	29.72	178.0, 294.5	14.61	207.6, 264.9
$C_{0,pbo}$	346.8	31.08	285.9, 407.7	22.26	303.2, 390.4

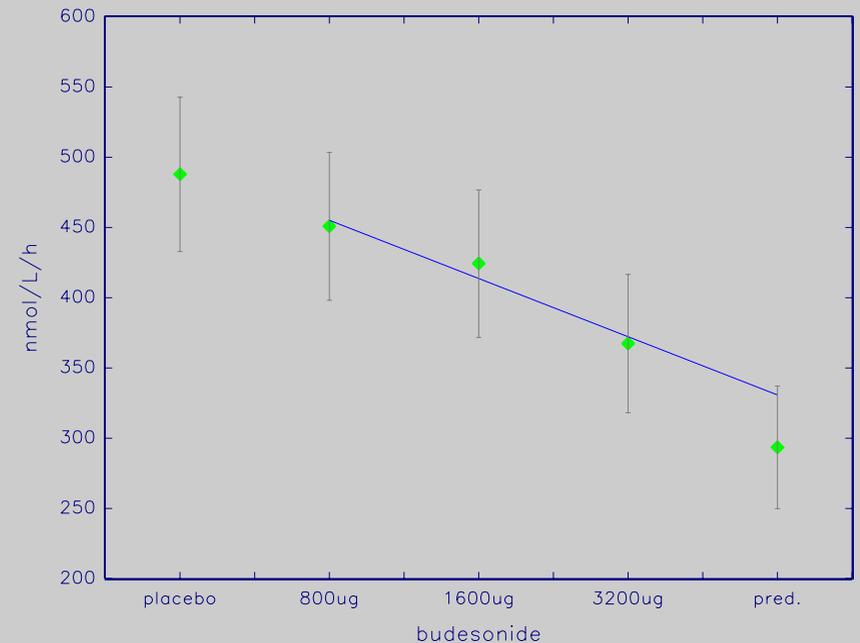


Ultimate objective: to find dose relation

Take the mean estimates of the I :s with its covariance matrix, and do a weighted linear regression for the three Pulmicort doses.

Then we estimate the dose of Pulmicort dose that, on average, gives the same effect as that of prednisolone 10 mg O.D. Confidence intervals are obtained using Fiellers method.

dose	95% confidence limits
11935	4390, 915940



Onset of action of peak flow

Background: To demonstrate clinical efficacy of an inhaled GCS requires larger studies, in which patients, at home, measure or assess and record into diary cards, among other things, peak expiratory flow (PEF) in the morning and evening.

Objective: To demonstrate the clinical efficacy of a new GCS, RPL in short.

Study Design: A parallel group study with two groups, RPL 100 μg b.i.d. and Placebo. Double-blind and double dummy. Two weeks baseline followed by four weeks of treatment. Diary cards – about 35 PEF morning measurements per patient.

Traditional analysis Compute two variables

1. Mean of all baseline measurement, to be used as baseline
2. Mean of last 14 days on treatment, to be used as effect variable



Result of traditional analysis:

	estimate	standard error	95% confidence limits
Effect	47.95	8.084	31.89, 64.01
INTERCEPT	440.6	5.59	429.5, 451.8

SD = 39 L/min (analysis uses baseline as covariate)

- Ignores half the data on treatment.
- If we took the average over the full treatment period, we would mix steady state and onset of action, leading to deflation of effect size.
- Single measurements from early withdrawals weight equal to completers data. Problem?

We attempt a mixed effects model which

- Uses all data, but differentiates between onset of action and steady state.
- Is as close as possible in philosophy to the traditional analysis
- Focuses on population mean behavior - first order approximation



NONMEM model for diary card data

- To model the mean placebo behavior, use a simple quadratic polynomial in time:

$$E(y_{pbo}) = E_0 + b_1t + b_2t^2 + b_3t^2$$

where t is number of days since randomization.

- To model the **effect**, use the function

$$f(b, t) = E_{max}(1 - e^{-kt}),$$

which starts at zero and asymptotically reaches E_{max} with a rate determined by k (which is $\ln 2/\text{time to reach 50\% of effect}$).

- The effect of RPL is modelled by

$$E(y_{rpl}) = E(y_{pbo}) + f(b, t).$$

- We assume only E_0 and E_{max} to be random parameters. Thus, again, a quasi-linear problem. k is parametrized in $\ln k$.



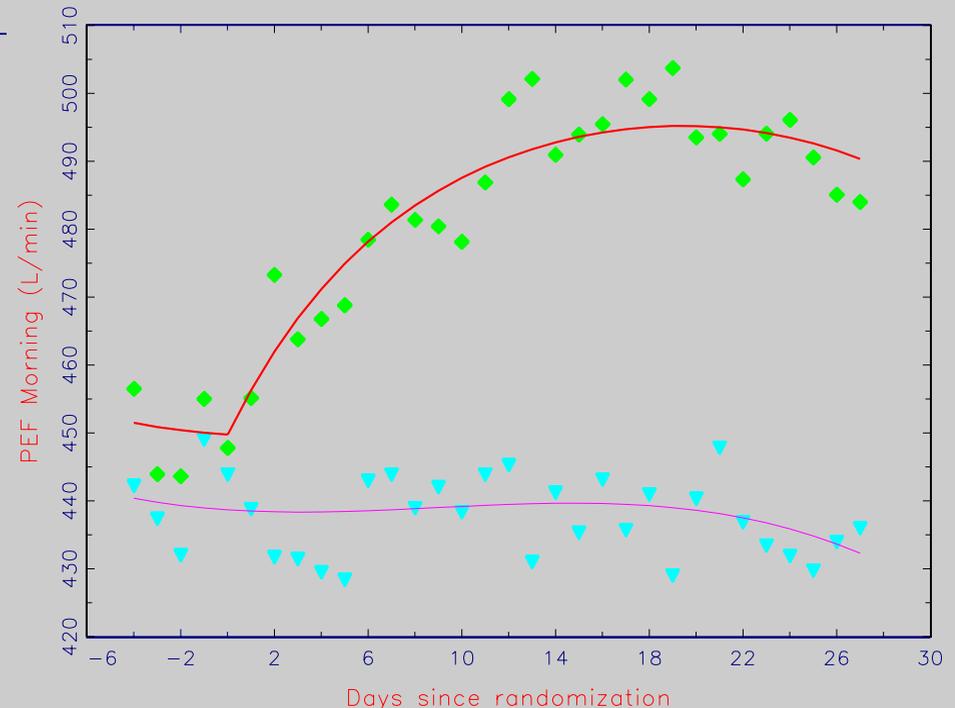
NONMEM results for diary card data However, we estimate using the first order, GEE-method, because we want total focus on the population mean curve!

	estimate	standard error	95% confidence limits
E_0	443.8	9.863	424.5, 463.2
E_{max}	47.78	8.89	30.36, 65.21
$\ln(k)$	-1.886	0.3621	-2.596, -1.177
b_1	-0.2264	0.5548	-1.314, 0.8609
b_2	0.0442	0.05239	-0.05848, 0.1469
b_3	-0.001652	0.001422	-0.004438, 0.001135

Within subject SD = 33.9

Random effects SD: $E_0 = 94.6$, $E_{max} = 41.7$

SD in ANOVA $\approx \sqrt{41.7^2 + 33.9^2/14} = 42.7$



Comparison to traditional analysis

- Almost the same result!
- Robust Estimator that is used, is slightly unbiased. Can be improved on! Nonlinear confidence intervals are slightly shorter in this case.
- Important extra information: The time to half the maximal effect (of the population mean curve) is estimated to 4.6 days, with a 95% confidence interval of (2.2, 9.3) days.



Dose response in the presence of a partial agonist

Background: A drug stimulating a receptor is called an agonist (to the receptor). If the response it gives is not maximal, it is called a partial agonist. In asthma: terbutaline is a full but short-acting agonist for the β_2 -receptor (leading to dilatation of airways in the lung), whereas salmeterol is a long-acting, partial agonist. It is also slow in terms of onset of action.

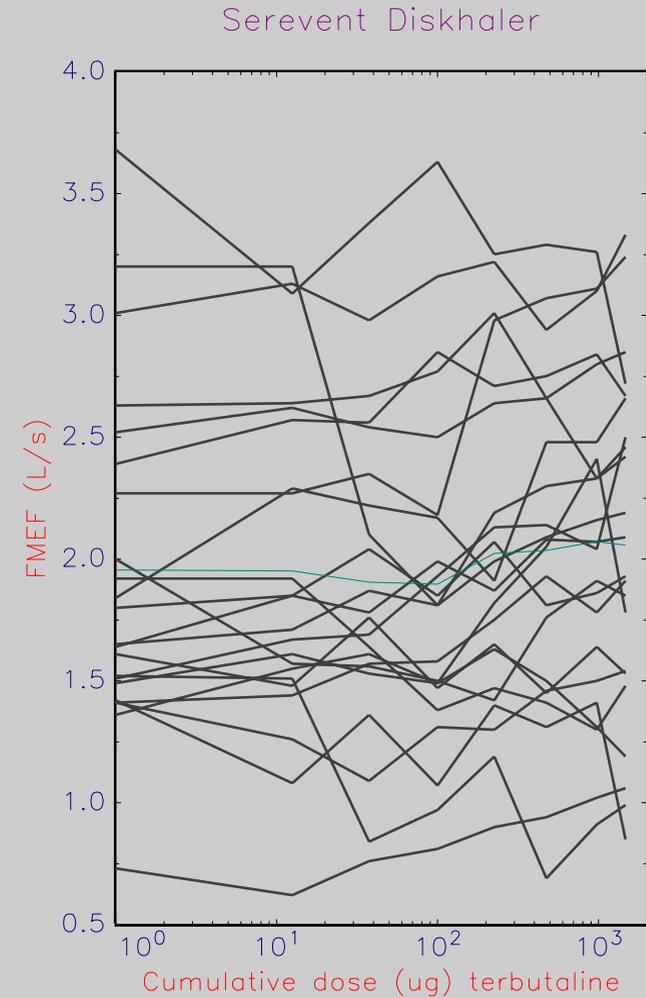
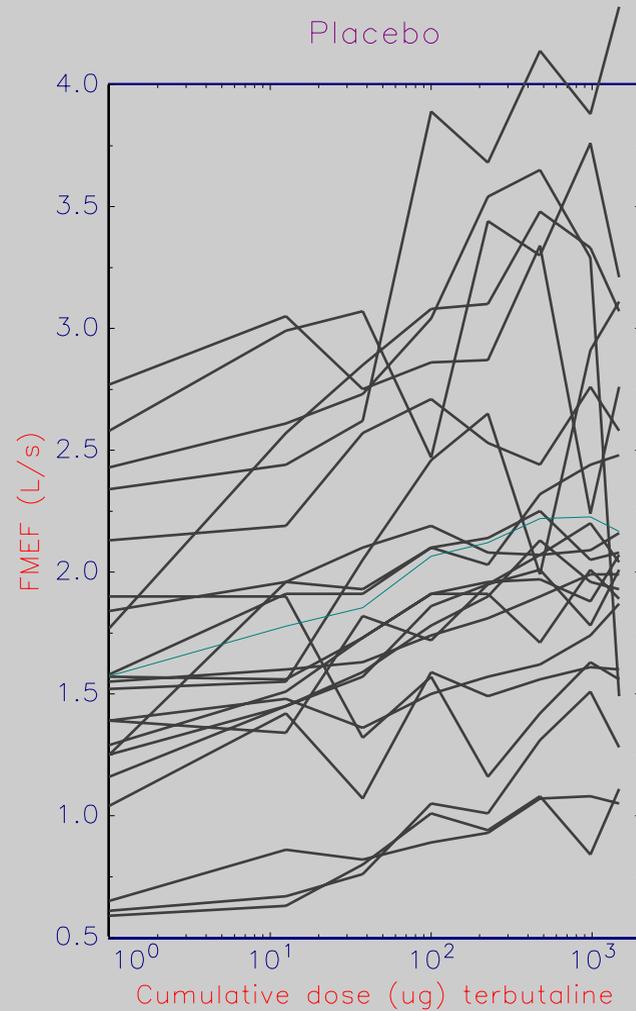
Objective: To see if the response to terbutaline is compromised if the patient is on regular treatment with salmeterol.

Study Design: A (double-blind, double-dummy) 2-period crossover study: one period with regular treatment with salmeterol, one with placebo. After each treatment period the dose response to terbutaline is investigated.



Individual data and mean data

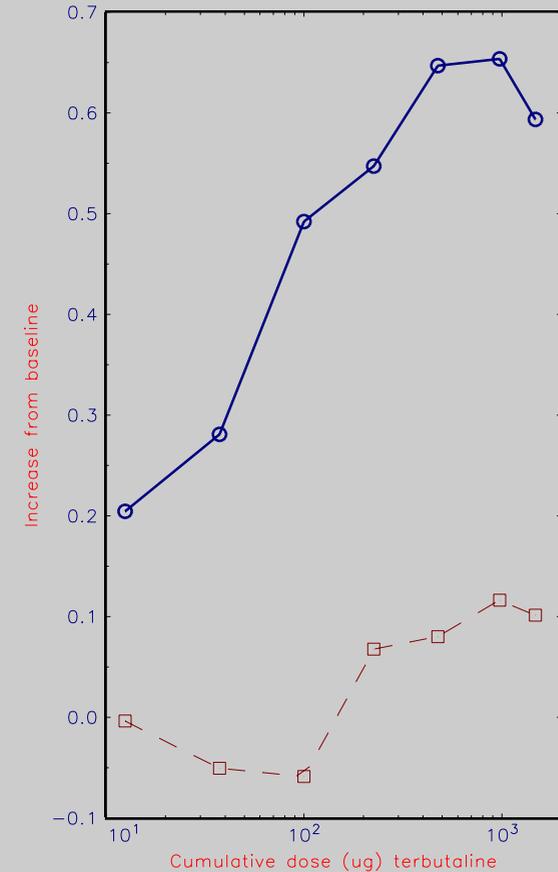
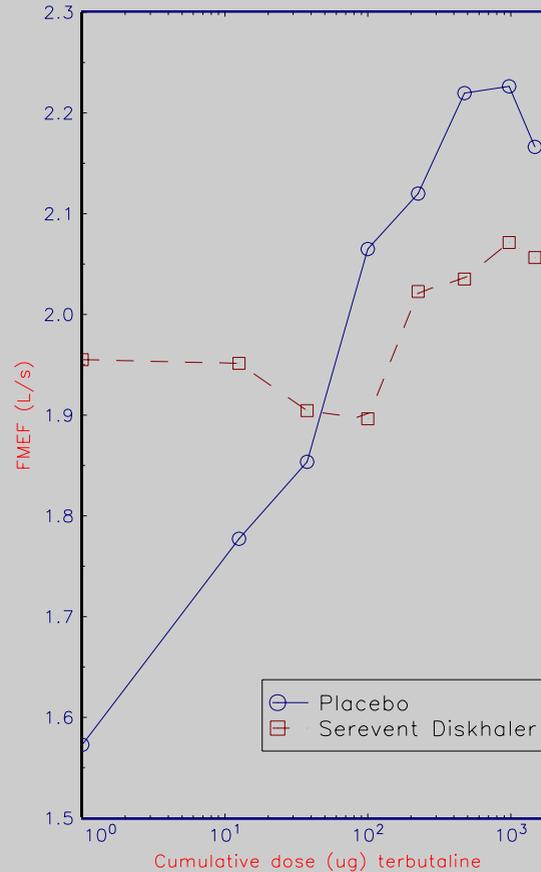
The figure to the right shows both individual data and mean data for the dose response test after pre-treatment with investigational drug.



The purpose of the analysis....

... is to see if we can demonstrate that more terbutaline is needed to achieve a given response, up and above the baseline response of the partial agonist.

The clinical implication could be that the baseline treatment with the partial agonist compromises your ability to relieve an acute attack that broke through. (If there are such attacks is another matter.)



Regression model for dose response experiment For the placebo curve we use a classical sigmoidal dose-response curve (with Hill coefficient one):

$$E_P(D) = E_0 + E_{max} \frac{D}{ED_{50} + D}.$$

Here dose D is the cumulative dose. For the salmeterol treatment we use

$$E_S(D) = E_0 + E_{max} \frac{D + \alpha ED_{50}(P - 1)}{ED_{50}P + D},$$

which can be rewritten

$$E_S(D) = (E_0 + \gamma E_{max}) + (1 - \gamma) E_{max} \frac{D}{ED_{50}P + D}, \quad \gamma = \alpha \left(1 - \frac{1}{P}\right).$$



Model rational

- The model for salmeterol increases monotonically from the baseline $E_0 + \gamma E_{max}$ to E_{max} . The parameter P measures any possible shift in dose required, i.e. how much the full agonist and the partial agonist *interact*. P is the parameter of interest!
- if $P > 1$, a high enough response requires more dose of the full agonist when given together with the partial agonist, than without it:
(to get the response $E_0 + \alpha E_{max}$ requires dose
 - $D = \alpha ED_{50}/(1 - \alpha)$ without, but
 - $D = (\alpha - \gamma)ED_{50}P/(1 - \alpha)$ with the partial agonist.The latter is larger when $P > \alpha/(\alpha - \gamma)$.)
- The first form of the salmeterol model can be motivated from the chemical steady state:



Analysis

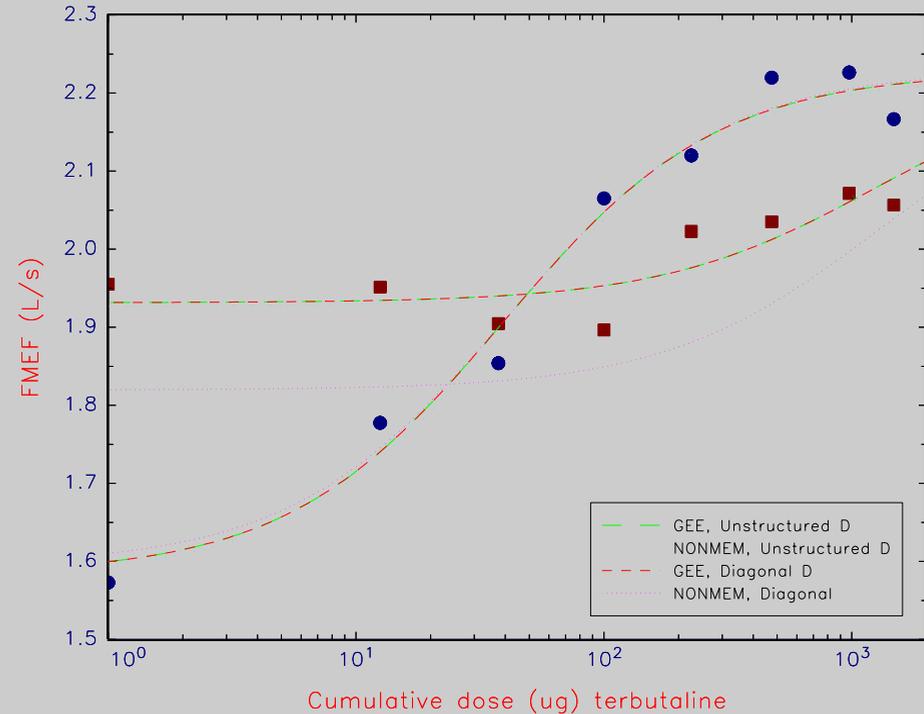
We analyze these data under four different first order models:

Two estimation methods:

1. GEE estimated
2. NONMEM

Two models for D :

1. Unconstrained
2. Diagonal



We see that three agree well, but the NONMEM with diagonal D differ from them.

A reasonable explanation is that

- The variance model is miss-specified
- NONMEM cares equally much about fitting to mean as to covariance

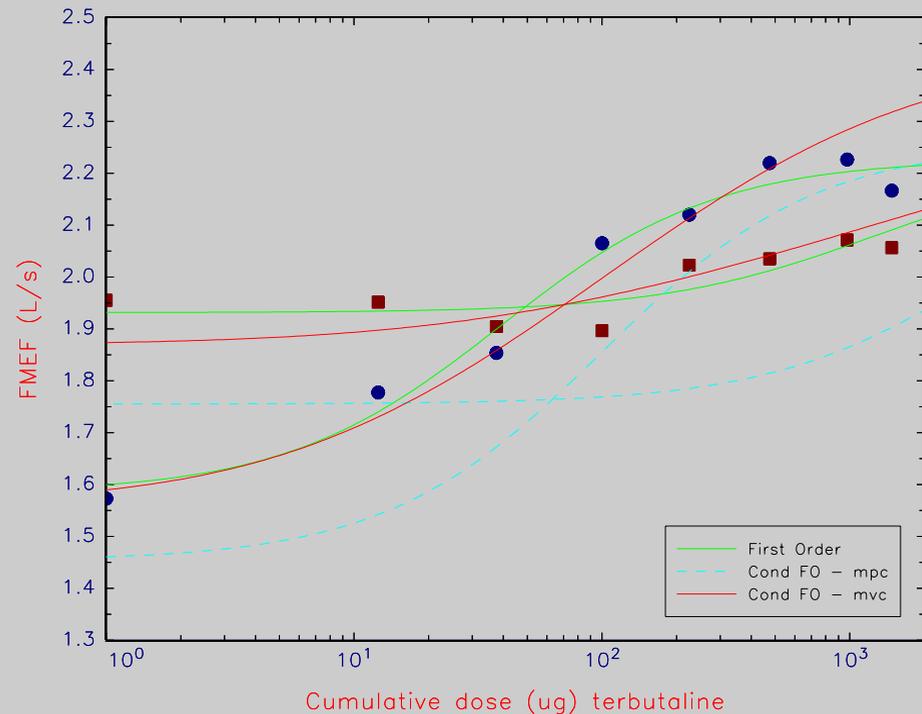


Conditional first order vs first order

We compare the GEE approach with diagonal D .

First Order Parameter	estimate	95% confidence limits
ED_{50}	38.92	23.1, 65.59
P	31.54	2.53, 393.2
$\frac{\alpha}{1-\alpha}$	1.259	0.6301, 2.516
E_{max}	0.6441	0.4659, 0.8904
E_0	1.584	1.354, 1.852

Conditional First Order Parameter	estimate	95% confidence limits
ED_{50}	100.1	42.66, 234.9
P	34.53	5.162, 231
$\frac{\alpha}{1-\alpha}$	0.6304	0.3118, 1.275
E_{max}	0.8056	0.6514, 0.9964
E_0	1.452	1.231, 1.713



The conclusion is that $P > 1$



Lecture 6: Generalized linear mixed effects models

In which we discuss another type of nonlinear mixed effects models, those based on generalized linear models.



The generalized linear model (GLM)

A distribution from the exponential family can be written in terms of its *natural parameter* θ as

$$p(y|\theta, \phi) = \exp\left(\frac{y^t\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

The basic property of this is that $\mu = E(y) = b'(\theta)$ and $V(y) = \phi b''(\theta)$. The first of these equations defines $\theta = \theta(\mu)$.

For a generalized linear model we assume that there is a link function $g(t)$ such that $g(\mu) = A\beta$.

Examples

- Logistic regression (binomial regression) and relatives, like Probit, for binary response
- Poisson regression for counts
- Multinomials, extending the binary case
- Gaussian, Gamma, Inverse Gaussian etc...



Poisson Regression

For a Poisson model we have

$$p(y) = e^{-m} m^y / y! = \exp(y \ln m - m - \ln y!) \implies \theta = \ln m, \quad b(t) = e^t \quad \phi = 1.$$

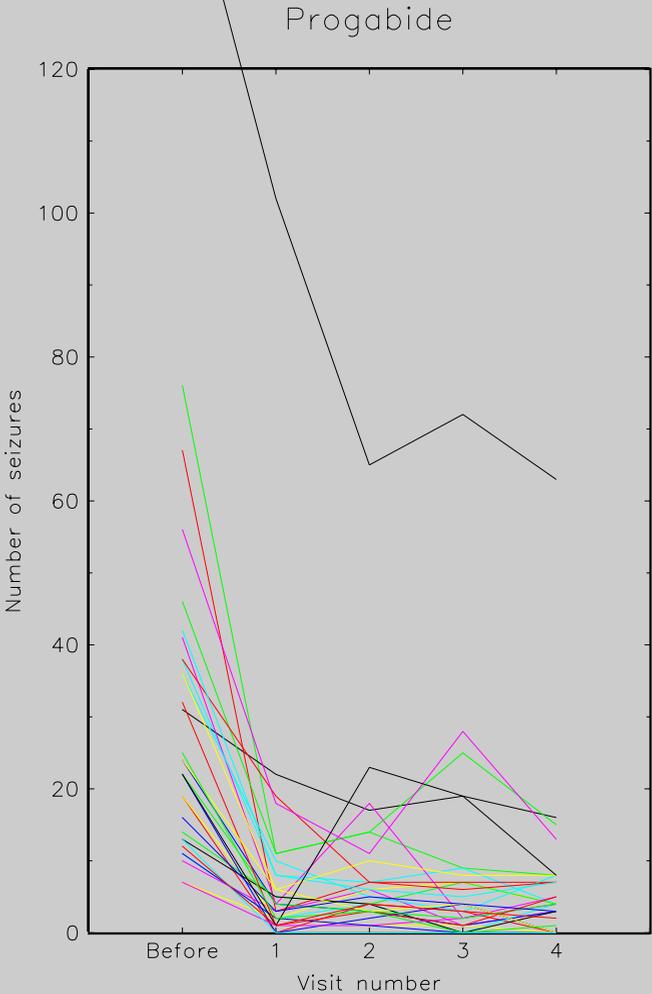
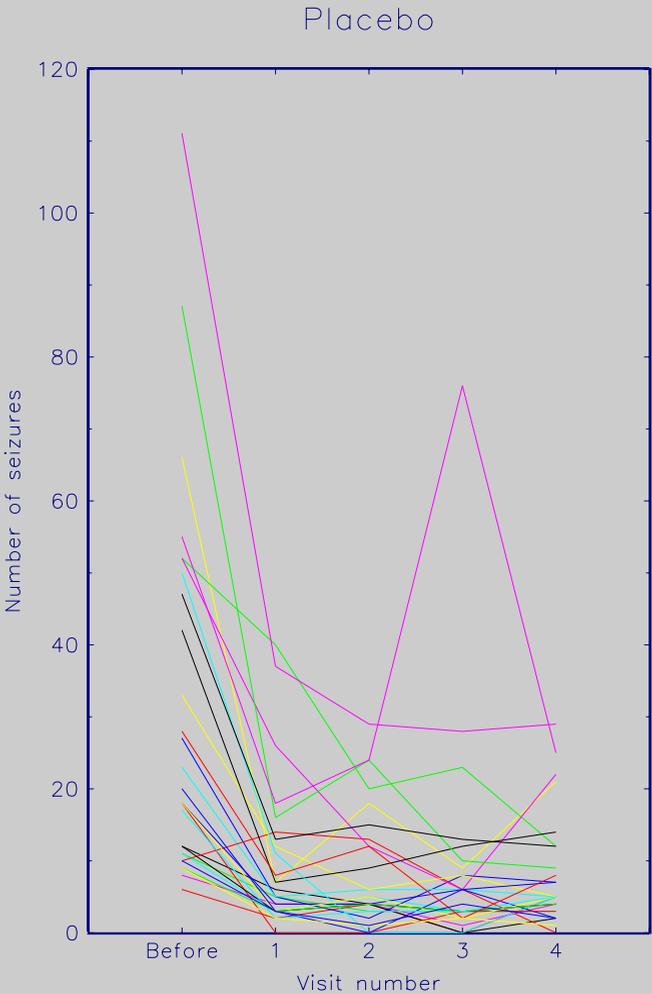
- Fifty-nine epileptics suffering from partial seizures were randomized to receive either progabide or placebo.
- The number of seizures in each of four two-week intervals were recorded, together with the number of seizures during an eight-week baseline period.
- To analyse this we use the following model:

$$\ln \mu_{ij} = \ln t_{ij} + \theta_0 + \theta_1 \text{Time} + \theta_2 \text{Treat} + \theta_3 \text{Time} * \text{Treat},$$

where $t_{ij} = 8$ if $j = 0$ and $t_{ij} = 2$ if $j > 0$ and where $Time$ is 0 at baseline and 1 else and $Treat$ is 1 if the treatment is progabide and 0 if it is placebo.



Here is a plot of the individual data



Result of Poisson Regression

Source	Deviance	df	P
Null	3581.8	294	
Model	6.3849	3	0.0943
Residual	3575.4	291	

Estimated (Pearson) Dispersion parameter: 19.68 (It is set to 1)

Parameter	estimate	standard error	95% confidence limits
Intercept	1.348	0.03406	1.281, 1.414
Time	0.1118	0.04688	0.01996, 0.2037
Treat	0.02753	0.04668	-0.06397, 0.119
Time*Treat	-0.1047	0.06503	-0.2322, 0.02273

Overdispersion indicates (substantial) correlation between observation – natural since we study different individuals, with different seizure rates, repeatedly.



Clusters of GLMs - the marginal approach (GEEs)

Overdispersion leads to faulty standard errors for coefficient estimates. One way to remedy that is to use the robust estimator:

Parameter	estimate	standard error	95% confidence limits
Intercept	1.348	0.1574	1.039, 1.656
Time	0.1118	0.1159	-0.1154, 0.3391
Treat	0.02753	0.2218	-0.4072, 0.4622
Treat*Time	-0.1047	0.2134	-0.5231, 0.3136

This is essentially the GEE technique (the marginal approach to analysis of this data). The GEE technique is built around two concepts:

- Correct mean function specification, but covariance structure a guess
- Compensate the wrong covariance structure by using the robust estimator for parameter covariance.



The generalized linear mixed effects model

Another way to account for inter-subject variability is to let some parameters depend on subject. Let

$$\beta_i = A_i\beta + C_i\xi, \quad \xi \in N(0, D).$$

Write

$$\mu_i(\beta, \xi) = g^{-1}(A_i\beta + C_i\xi)$$

and we get

$$L_m(\beta, D, \phi) = \prod_{i=1}^N \int e^{-Q_i(\beta, D, \phi, \xi)/2} d\xi,$$

where

$$Q_i(\beta, D, \phi, \xi) = -2 \ln p(y_i | \theta_i, \phi) + \xi^t D^{-1} \xi + \ln(\det 2\pi D).$$

This integral can be approach in complete analogy with the Gaussian case ($\theta_i = \theta_i(\beta, \phi)$ is the individual, natural parameter for a GLM).



In the Poisson case we get

$$\mu_i(\beta, \xi) = \exp(A_i\beta + C_i\xi)$$

which means that we can compute the mean value curve explicitly:

$$\mu_i(\beta, D) = (\det 2\pi D)^{-1/2} \int \mu_i(\beta, \xi) e^{-\xi^t D^{-1} \xi / 2} d\xi = \exp(A_i\beta + C_i D C_i^t / 2).$$

Concerning the likelihood, we have

$$\ln p(y_i | \theta_i) = y_i(A_i\beta + C_i\xi) - \exp(A_i\beta + C_i\xi) - \ln y_i!,$$

so the integral to compute becomes

$$L_m(\beta, D) \propto (\det 2\pi D)^{-1/2} \int \exp(y_i(A_i\beta + C_i\xi) - \exp(A_i\beta + C_i\xi) - \xi^t D^{-1} \xi / 2) d\xi$$



Random intercept only

As a first GLMEM, consider the case with only the intercept being random. Corresponds to each subject having an individual basic seizure intensity.

Result:

Parameter	estimate	standard error	95% confidence limits
Intercept	1.034	0.1888	0.6634, 1.404
Time	0.112	0.04689	0.02009, 0.2039
Treat	-0.02209	0.2906	-0.5917, 0.5475
Treat*Time	-0.1051	0.06481	-0.2321, 0.02197

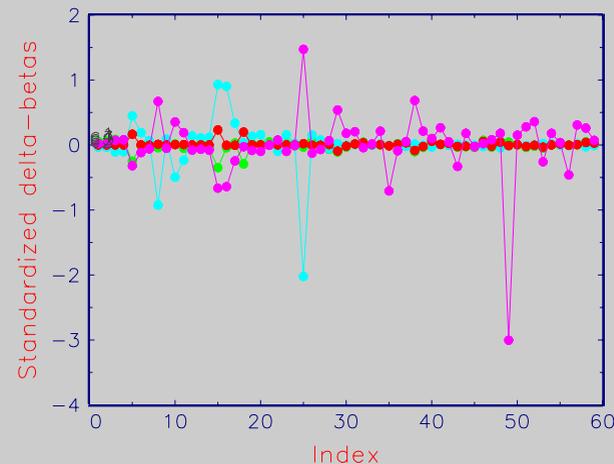
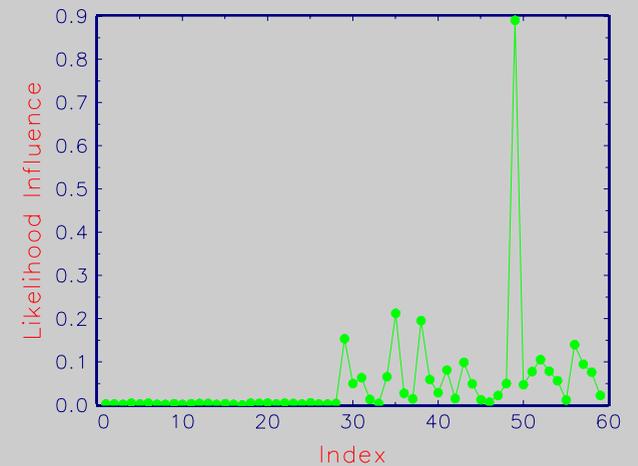
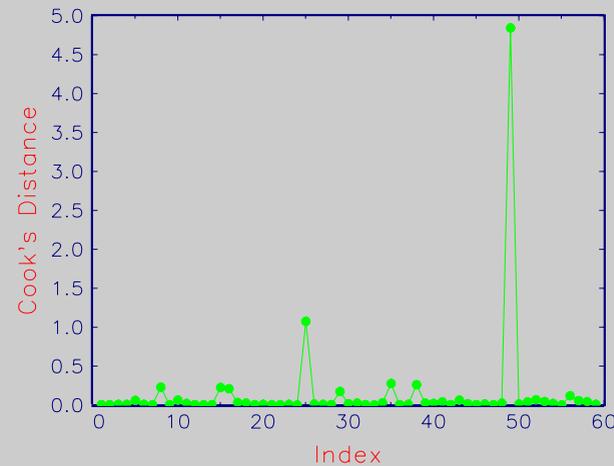
The random SD was estimated to 0.78

(note that $1.35 \approx 1.03 + 0.78^2/2$)



To the right is a diagnostic plot of this data

We see that subject 49 has a heavy influence on the result. Excluding him, we get a $\text{Treat} \times \text{Time}$ of the order -0.31 and statistically significant! That subject has an unusual set of counts.



Random intercept and time

Next we consider the case where both the intercept and time parameters are random. Thus each individual have a basic, individual seizure intensity, which may differ between baseline period and treatment period.

Result:

Parameter	estimate	standard error	95% confidence limits
Intercept	1.071	0.1854	0.7072, 1.434
Time	0.001038	0.1113	-0.2172, 0.2193
Treat	0.05131	0.2555	-0.4495, 0.5521
Treat*Time	-0.3088	0.1482	-0.5992,-0.01838

$$D = \begin{pmatrix} 0.1221 & -0.2926 \\ -0.2926 & 1.191 \end{pmatrix}$$

We see that Treat*Time is significant, implying an effect of the treatment!

